

909/897188

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 88.6045 Seconds
(without alignments)
4449.477 Million cell updates/sec

Title: SEQ1-A
Perfect score: 5835
Sequence: 1 MAWKTLPIYLLLSVFEVQ.....DMYLPVNPQGIINPMLS 1099

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : A_Geneseq_23Sep04:*
- 1: Geneseqp1980s:*
 - 2: Geneseqp1990s:*
 - 3: Geneseqp2000s:*
 - 4: Geneseqp2001s:*
 - 5: Geneseqp2002s:*
 - 6: Geneseqp2003as:*
 - 7: Geneseqp2003bs:*
 - 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5820.9	99.8	1299	4	AAM24322 Human EST
2	5820.9	99.8	1404	2	AAR26049 MSF precu
3	5820.9	99.8	1404	4	AAB29773 Human meg
4	5820.9	99.8	1404	4	AAB60568 Human meg
5	5820.9	99.8	1404	8	ADM98014 Human meg
6	5820.9	99.8	1415	4	AAU32262 Novel hum
7	5820.9	99.8	1404	7	ADK65839 Angiogene
8	5808.7	96.1	1320	7	ADK65819 Angiogene
9	5353.7	91.8	1270	8	ADK67911 Human ext
10	5339.6	91.5	1311	8	ADK67912 Human ext
11	2920	50.0	546	4	ABU53252 Human tes
12	2888	49.5	551	4	ABU53253 Human tes
13	2880.1	49.4	902	4	ABU53254 Human tes
14	2757	47.2	513	4	ABU53254 Human tes
15	1954.5	33.5	452	2	AAO18834 3' cartil
16	1409.9	24.2	538	5	AAO18834 3' cartil
17	1401.6	24.0	292	5	AAU11261 Human HAP
18	1354.4	23.2	5179	4	AAM24516 C899P pre
19	1354.4	23.2	5179	6	ABP55365 Human col
20	1354.4	23.2	5179	6	ABO07258 Human p53
21	1354.4	23.2	5179	7	ADDA48091 Human Pro
22	1354.4	23.2	5179	7	ADDA4998 Human Pro
23	1343.3	19.6	1684	2	AAW43106 C. thermo
24	1131.8	19.4	8991	6	ABU08487 S. pneumo
25	1099.6	18.8	717	4	ABU53144 Human tes

ALIGNMENTS

RESULT 1
AAM24322
ID AAM24322 standard; protein; 1299 AA.
XX
AC AAM24322;
XX

DT 12-OCT-2001 (first entry)
XX
DE Human EST encoded protein SEQ ID NO: 1847.
XX

KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;
KW Gene therapy; nutrition.
XX

OS Homo sapiens.
XX
XX WO200154477-A2.
XX
PD 02-AUG-2001.
XX

PF 25-JAN-2001; 2001WO-US002687.
XX
PR 25-JAN-2000; 2000US-00491404.
PR 17-JUL-2000; 2000US-00617746.
PR 03-AUG-2000; 2000US-00631451.
PR 15-SEP-2000; 2000US-00663870.
XX

PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX

DR WPI: 2001-476164/51.
DR N-PSDB; AAH98981.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use.
XX

XX Claim 20; Page 1198-1201; 12/5pp; English.
PS
XX The present invention provides the protein and coding sequences of novel
XX proteins from a variety of organisms, including human, dog, cat, horse,
XX cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
XX urchin and tomato. These were derived from expressed sequence tags (ESTs)
XX from the organism of interest. They can be used in diagnostics,
XX forensics, gene mapping, identification of mutations, to assess
XX biodiversity and for nutritional purposes. The present sequence is a

Abu53255 Human tes
Abb69806 Drosophil
Aag38942 Arabidops
Abb61364 Drosophil
Abb71434 Drosophil
Adi23265 Human MUC
Abb60403 Drosophil
Abb43908 MUC5B par
Abu53141 Human tes
Abu53143 Human tes
Abu53155 Human tes
Adi21202 Novel hum
Adi39110 Cancer/an
Abu53150 Human tes
Abu53149 Human tes
Abu53151 Human tes
Abu53148 Human tes
Abu53147 Human tes
Abu53146 Human tes

26 1097 18.8 214 4 ABU53255
27 1087.6 18.6 1795 4 ABB69806
28 1052.6 18.0 763 3 AAG38942
29 1044.2 17.9 1049 4 ABB61364
30 1034.5 17.7 2284 4 ABB71434
31 998.1 17.1 5703 8 ADI23265
32 996.1 17.1 2112 4 ABB60403
33 980.5 17.0 4315 5 ABB43908
34 987.9 16.9 770 4 ABU53141
35 981 16.8 778 4 ABU53143
36 974.4 16.7 692 4 ABU53155
37 973.4 16.7 1538 7 ADI21202
38 972.9 16.7 1460 7 ADI39110
39 964.1 16.5 717 4 ABU53150
40 964.1 16.5 717 4 ABU53149
41 964.1 16.5 717 4 ABU53151
42 964.1 16.5 717 4 ABU53148
43 964.1 16.5 717 4 ABU53147
44 964.1 16.5 717 4 ABU53146
45 964.1 16.5 717 4 ABU53146

Claim 1, 2 and 3; Fig 1; 87pp; English.

The sequence given is a full length translation from the megakaryocyte stimulating factor (MSF) precursor. The sequence covered by exons I, III and IV encodes megakaryocyte stimulating factor (MSP). This sequence is modified by the addition of an N-terminal sequence encoding a secretory leader, an initiating methionine preceding exon II and a terminating codon following exon IV. The cDNA sequence given contains sequences derived from human megakaryocyte colony stimulating factor (meg-CSF). Exon I contains the initiating methionine, and encodes a classical mammalian protein secretion signal sequence. The sequence encoding the original meg-CSF includes exons II-IV and is thought to terminate in the region between amino acid residues 134 - 147. The primary transcript of this gene may be cleaved in different ways to yield a family of mRNA's each encoding a different MSF protein. Exons V and VI are thought to be related to the activity of the factor and are also implicated in the stability, folding and processing of the molecule. These exons are also thought to play a role in the observed synergy of MSF with other cytokines. Exons V - XII are believed to be implicated in the processing or folding of the appropriate structure of the resulting factor, i.e. one or more of these exons may contain sequences which direct proteolytic cleavage, adhesion, organisation of the cellular matrix or extracellular matrix processing. Both naturally occurring and non-naturally occurring MSP's may be characterised by various combinations of alternatively spliced exons from this sequence, with the exons spliced together in differing orders to form different members of the MSF family. (Updated on 25-MAR-2003 to correct PN field.)

XX XX Sequence 1404 AA;
Query Match 99.8%; Score 5820.9; DB 2; Length 1404;
Best Local Similarity 96.4%; Pred. No. 6.1e-157;
Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQQVSSQ----- 25
DB 1 MAWKTLPIYLLLSVFIQQVSSQLSSCAGRCGEYSRDTNCNDYNCOHYNECCPDF 60
QY 26 -----ELSCKGRCFESPERGECDDAQCCKYDKCPDYESFCAEVHNPTSPSSKKAP 79
DB 61 KRVTALSLCKGRCFESPERGECDDAQCCKYDKCPDYESFCAEVHNPTSPSSKKAP 120
QY 80 PPSGASQTIKSTTKRSKPKNKXTKKVIESEBITTEHSVSENQESSSSSSSSSSTIHW 139
DB 121 PPSGASQTIKSTTKRSKPFPNKKXTKVIESEBITTEHSVSENQESSSSSSSSSSTIHW 180
QY 140 KIKSKNSAANRELQKLKVXDNKNRTKKXTPPKPPVVVDAGSLDNGDPKVTITPDST 199
DB 181 KIKSKNSAANRELOKKLVXDNKNRTKKXTPPKPPVVVDAGSLDNGDPKVTITPDST 240
QY 200 TQHNKVSPTSKITTAACPINRPSPLPNSDTSKETSLTVNKETTVEKTETTTNNKOTSTDG 259
DB 241 TQHNKVSPTSKITTAACPINRPSPLPNSDTSKETSLTVNKETTVEKTETTTNNKOTSTDG 300
QY 260 KEKITSAKETSQIESAKDLAPTSLVKLAKPTPKAEITTKGPALTTPKEPPTTPEKAS 319
DB 301 KEKITSAKETSQIESAKDLAPTSLVKLAKPTPKAEITTKGPALTTPKEPPTTPEKAS 360
QY 320 TTPEKPTPTTIKSAPTTTPKEPAATTTKSAPTTTPKEPAATTTTPKEPATTTTXEP 379
DB 361 TTPEKPTPTTIKSAPTTTPKEPAATTTKSAPTTTPKEPAATTTTPKEPATTTTXEP 420
QY 380 APTTTKSAPTTTPKEPAATTPPKAPATTTPKEPATTPKEPTTTPKEPATTPKAP 439
DB 421 APTTTKSAPTTTPKEPAATTPPKAPATTTPKEPATTPKSPPTTTPKEPATTTTPKEPAT 480
QY 440 EPAPTA PKKAPATTTPKEPAATTPPKAPATTTKBPGSTTPKEPAATTTKSAPTTTKEPAT 499
DB 481 EPAPTA PKKAPATTTPKEPAATTPKEPAATTTKEPSPTTPKEPATTTKSAPTTTKEPAT 540
QY 500 TTKSAPTTTPKEPSPTTTKEPATTPKEPATTPKKAPATTTPKEPATTPPKEPATTTTKP 559

XX PS Claim 3; Page 7; 47pp; English.

XX CC The invention relates to a human tribonection which is a product of

CC alternative splicing of the human MSF (megakaryocyte stimulating factor)

CC gene. The tribonection has at least one O-linked oligosaccharide

CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats

CC of a motif having at least 50% identity to the sequence KSPAPPT

CC (AAB29774). The invention also relates to a nucleic acid encoding a human

CC MSF-derived tribonection; a biocompatible composition comprising a human

CC tribonection for inhibiting tissue adhesion formation; and a method of

CC diagnosing osteoarthritis or a predisposition to osteoarthritis by

CC measuring the amount of MSF or its fragment in a biological sample of a

CC mammal, wherein an increased amount of MSF compared to a control

CC indicates the presence of or predisposition to developing osteoarthritis.

CC The tribonection and DNA encoding it are useful in the treatment of

CC osteoarthritis, where they may be used for lubricating mammalian joints,

CC such as articulating joints of humans, dogs or horses. The tribonection,

CC when formulated as a membrane, foam, gel or fibre, is useful for

CC inhibiting adhesion between two surfaces such as the injured tissues of a

CC mammal, where the injury is caused by a surgical insertion or trauma, or

CC an artificial device e.g., an orthopaedic implant. In particular, one of

CC the surfaces is pericardial tissue. DNA encoding a tribonection may be

CC used in gene therapy. The present sequence represents human MSF

XX SQ Sequence 1404 AA;

Query Match 99.8%; Score 5820.9; DB 4; Length 1404;

Best Local Similarity 96.4%; Pred. No. 6.1e-157;

Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25

DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYHCQHYMECCPDF 60

QY 26 -----ELSCKGRCPESFERGECDDCAQCKYKCCPDYBSFCAEVHNPTSPSSKAP 79

DB 61 KRVCYTAELSCKGRCPESFERGECDDCAQCKYKCCPDYBSFCAEVHNPTSPSSKAP 120

QY 80 PPSGASQIKSTTKBSPPNKKTKKVIIEBETEEHSVSENQESSSSSSSSSSSTIW 139

DB 121 PPSGASQIKSTTKBSPPNKKTKKVIIEBETEEHSVSENQESSSSSSSSSSSTIW 180

QY 140 KIKSSKNSAANRELQKLVKXDNKNRTKKXPTPKPVVDAGSLDNGPKVTTPTST 139

DB 181 KIKSSKNSAANRELQKLVKXDNKNRTKKXPTPKPVVDAGSLDNGPKVTTPTST 240

QY 200 TQHNKVSTSPKITTAKPINRPPSPNSDTSKETSLSVKNKETTVEKTTTNKQTS DG 259

DB 241 TQHNKVSTSPKITTAKPINRPPSPNSDTSKETSLSVKNKETTVEKTTTNKQTS DG 300

QY 260 KEKTTISAKETOSIEKTSKADLAPTSKVLAKPTPKAETTKGPALTTKEPTTPKBPAS 319

DB 301 KEKTTISAKETOSIEKTSKADLAPTSKVLAKPTPKAETTKGPALTTKEPTTPKBPAS 360

QY 320 TTPKEPTTTIKSAPTTPKBPATTTKSAPTTTPKBPATTTTPKBPATTTTPKBPATTTTKEP 379

DB 361 TTPKEPTTTIKSAPTTPKBPATTTKSAPTTTPKBPATTTTPKBPATTTTPKBPATTTTKEP 420

QY 380 APTTTKSAPTTTPKBPATTTPKKRAPTTTPKBPATTTTPKBPATTTTPKBPATTTTPK 439

DB 421 APTTTKSAPTTTPKBPATTTPKKRAPTTTPKBPATTTTPKBPATTTTPKBPATTTTPK 480

QY 440 EPAPTAPKBPATTPKBPATTTTPKBPATTTTPKBPATTTTPKBPATTTTPKBPATTTTKEPAPT 499

DB 481 EPAPTAPKBPATTPKBPATTPKBPATTTTPKBPATTTTPKBPATTTTPKBPATTTTKEPAPT 540

QY 500 TTKSAPTTTPKBPSTTTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTTTPKBP 559

DB 541 TTKSAPTTTPKBPSTTTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTTTPKBP 600

QY 560 APTAPKEPAPTTTPKETAPTTPKLTPPTPEKLAPTTPEKAPTTPEELAPTTPEEPTPTT 619

DB 601 APTAPKEPAPTTTPKETAPTTPKLTPPTPEKLAPTTPEKAPTTPEELAPTTPEEPTPTT 660

QY 620 PEEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTTPKGTAPTTLKEP 679

DB 661 PEEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTTPKGTAPTTLKEP 720

QY 680 APTTPKPAKELAPTTTKEPTSTSDKRAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTPKG 739

DB 721 APTTPKPAKELAPTTTKEPTSTSDKRAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTPKG 780

QY 740 TAPTTLKEPAPTTTPKPAKELAPTTTKEPTSTSDKRAPTTTPKGTAPTTTPKEPAPTTTPK 799

DB 781 TAPTTLKEPAPTTTPKPAKELAPTTTKEPTSTSDKRAPTTTPKGTAPTTTPKEPAPTTTPK 840

QY 800 KPAPTTTPETPPPTTSEVSTPTTTTKEPTTHKSDPSTPSELSAEPPTKALENSPKPEGVPT 859

DB 841 KPAPTTTPETPPPTTSEVSTPTTTTKEPTTHKSDPSTPSELSAEPPTKALENSPKPEGVPT 900

QY 860 TKTPAATKPEMTTAKDKITERDLRTPPTTTTAAAPKMTKETATTTTEKTTESKITATTTGV 919

DB 901 TKTPAATKPEMTTAKDKITERDLRTPPTTTTAAAPKMTKETATTTTEKTTESKITATTTGV 960

QY 920 TSTTTQDTPPFKLTTLTKTTLAPKVTTKKTTTTEIMNKPEETAKPKDRATNSKATTPK 979

DB 961 TSTTTQDTPPFKLTTLTKTTLAPKVTTKKTTTTEIMNKPEETAKPKDRATNSKATTPK 1020

QY 980 POKPTKAPKPTSTKKPKTMPRVKPKTTPTRKMTSTMPKLTSTMPKLTSTMPKLTSTMPKLTSTMPK 1039

DB 1021 POKPTKAPKPTSTKKPKTMPRVKPKTTPTRKMTSTMPKLTSTMPKLTSTMPKLTSTMPKLTSTMPK 1080

QY 1040 QTPNSKLVEVNPKESEDAGAEGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1099

DB 1081 QTPNSKLVEVNPKESEDAGAEGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1140

RESULT 4

AAB60568

ID AAB60568 standard; protein; 1404 AA.

XX AC AAB60568;

XX DT 27-APR-2001 (first entry)

XX DE Human megakaryocyte stimulating factor (MSF, CACP).

XX KW Human; CACP protein; camptodactyly-arthropathy-coxa vara-pericarditis;

XX KW MSF; megakaryocyte stimulating factor; synovial lubricant;

XX KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;

XX KW antiarthritic.

XX OS Homo sapiens.

XX PN WO200107068-A1.

XX PD 01-FEB-2001.

XX PF 21-JUL-2000; 2000WO-US0200002.

XX PR 23-JUL-1999; 98US-0145328P.

XX PR 19-JUL-2000; 2000US-00145328.

XX PA (UYCA-) UNIV CASE WESTERN RESERVE.

XX PI Warman ML;

XX DR WPI; 2001-192721/18.

XX PT New composition comprising the camptodactyly-arthropathy-coxa vara-

XX PT pericarditis protein in combination with an anesthetic, useful for

XX PT treating osteoarthritis, or as lubricants of tissue and joints.

XX PS Example 1; Page; 34pp; English.

XX

Qy	680	APTTPKKPAPKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPKPAPPTTKEPAPTTPKG	739
Db	721	APTTPKKPAPKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPKPAPPTTKEPAPTTPKG	780
Qy	740	TAPTTLKEPAPTPPKPAPKELAPTTTKEPTSTTSKDPAPTTPKETAPTTKEPAPTTPK	799
Db	781	TAPTTLKEPAPTPPKPAPKELAPTTTKEPTSTTSKDPAPTTPKETAPTTKEPAPTTPK	840
Qy	800	KPAPTTPEPTTPPTSEVSTPTTTKEPTTHKSPDESTPELSAEPKALENSPKPEGVPT	859
Db	841	KPAPTTPEPTTPPTSEVSTPTTTKEPTTHKSPDESTPELSAEPKALENSPKPEGVPT	900
Qy	860	TKTPAAATKPEMTTITAKDKTTERDLATTPETTTAAPKMTKETATTTTEKTTESKITATTQV	919
Db	901	TKTPAAATKPEMTTITAKDKTTERDLATTPETTTAAPKMTKETATTTTEKTTESKITATTQV	960
Qy	920	TSITTTODTTPFKITTLKTTTLAPKVTITTKTITTTTEIMNKPEETAKPKDRATNSKATTPK	979
Db	961	TSITTTODTTPFKITTLKTTTLAPKVTITTKTITTTTEIMNKPEETAKPKDRATNSKATTPK	1020
Qy	980	PQKPTKAPKKPTSTKKPKTMPVRVKPKITPTPRKMTSTMPELNPTSRIAEAMLOTTTRPN	1039
Db	1021	PQKPTKAPKKPTSTKKPKTMPVRVKPKITPTPRKMTSTMPELNPTSRIAEAMLOTTTRPN	1080
Qy	1040	QTPNSKLVEVNPKSEDAGGAEETHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS	1099
Db	1081	QTPNSKLVEVNPKSEDAGGAEETHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS	1140
RESULT 5			
ADM98014			
ID	ADM98014 standard; protein; 1404 AA.		
XX			
AC	ADM98014;		
XX			
DT	01-JUL-2004 (first entry)		
XX			
DE	Human megakaryocyte stimulating factor (MSF).		
XX			
KW	lubricating polypeptide; O-linked oligosaccharide; joint lubrication;		
KW	CAP; camptodactyl-arthropathy pericarditis; osteoarthritis; human;		
KW	megakaryocyte stimulating factor; MSF.		
XX			
OS	Homo sapiens.		
XX			
PN	US2004072741-A1.		
XX			
PD	15-APR-2004.		
XX			
PF	02-JUL-2001; 2001US-00897188.		
XX			
PR	23-APR-1999; 99US-00298570.		
PR	24-APR-2000; 2000US-00556246.		
XX			
PA	(JAYG/) JAY G D.		
XX			
PI	Jay Gp;		
XX			
DR	WPI; 2004-373948/35.		
DR	N-PSDB; ADM98015.		
XX			
PT	New tribonecin polypeptides and polynucleotides for lubricating joints		
PT	or other tissues to prevent or treat Camptodactyl-arthropathy-		
PT	pericarditis syndrome or osteoarthritis.		
XX			
PS	Claim 1; SEQ ID NO 1; 34pp; English.		
XX			
CC	The invention relates to a lubricating polypeptide and at least one O-		
CC	linked oligosaccharide. The composition and methods are useful for		
CC	lubricating joints or other tissues to prevent or treat camptodactyl-		
CC	arthopathy pericarditis (CAP) or osteoarthritis in mammals. The present		
CC	sequence represents the amino acid sequence of the human megakaryocyte		

```
CC stimulating factor (MGF).
XX sequence 1404 AA;
SQ

Query Match
Best Local Similarity 99.8%; Score 5820.9; DB 8; Length 1404;
Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVLIQVSSQ----- 25
DB 1 MAWKTLPIYLLLLSVFVLIQVSSQDLSSCAGRGEGYRDATCNCNDYHCQYMECCPDF 60

QY 26 -----ELSCGRCFFSFERGECDCDAQCKYDKCCPDYSCAEVHNFTSPSSKKAP 79
DB 61 KRVTAEIASCGRFCFSFERGECDCDAQCKYDKCCPDYSCAEVHNFTSPSSKKAP 120

QY 80 PPSGASQTIKSTTKRSPKPNKKTTKVIESEIIEHSVSNQSSSSSSSSSSSTIW 139
DB 121 PPSGASQTIKSTTKRSPKPNKKTTKVIESEIIEHSVSNQSSSSSSSSSSSTIW 180

QY 140 KIKSSKNSAANRELOKLLKVDKDKKQRTKKKTPKPPVVDGAGSLDNGDFKVTTPDTST 199
DB 181 KIKSSKNSAANRELOKLLKVDKDKKQRTKKKTPKPPVVDGAGSLDNGDFKVTTPDTST 240

QY 200 TOHNKUSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKTTTTNKQTSIDG 259
DB 241 TOHNKUSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKTTTTNKQTSIDG 300

QY 260 KKTTSKETSQIETKSADLAPTSKVLAKPTPKAETTTKGPALTPKPEPTTPKEPAS 319
DB 301 KKTTSKETSQIETKSADLAPTSKVLAKPTPKAETTTKGPALTPKPEPTTPKEPAS 360

QY 320 TTPKEPTPTTIKSAPTTPKEPAPTTKGAPTTPKEPAPTTKKEPAPTTTPKEPAPTTTKEP 379
DB 361 TTPKEPTPTTIKSAPTTPKEPAPTTKGAPTTPKEPAPTTKKEPAPTTTPKEPAPTTTKEP 420

QY 380 APTTTKSAPTTPKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTK 439
DB 421 APTTTKSAPTTPKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTK 480

QY 440 EPAPTAPKAPPTTPKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPT 499
DB 481 EPAPTAPKAPPTTPKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPT 540

QY 500 TTKSAPTTPKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEP 559
DB 541 TTKSAPTTPKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEP 600

QY 560 APTAPKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPT 619
DB 601 APTAPKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPT 660

QY 620 PEEAPPTPKAAAPNTPEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPT 679
DB 661 PEEAPPTPKAAAPNTPEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPT 720

QY 680 APTPKKAPKELAPTTTKEPTSTSDKAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPT 739
DB 721 APTPKKAPKELAPTTTKEPTSTSDKAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPT 780

QY 740 TAPTLKPEAPTTPKKAPKELAPTTTKEPTSTSDKAPTTKKEPAPTTKKEPAPTTKKEPAPT 799
DB 781 TAPTLKPEAPTTPKKAPKELAPTTTKEPTSTSDKAPTTKKEPAPTTKKEPAPTTKKEPAPT 840

QY 800 KPAPTTPPTPTTSEVSTPTTKEPTTIHKSDESTPELSAETPKALENSKEPCVPT 859
DB 841 KPAPTTPPTPTTSEVSTPTTKEPTTIHKSDESTPELSAETPKALENSKEPCVPT 900

QY 860 TKTPAATKPEMTTAKOKTTERDLRTPTPTTAAAPKMTKATATTEKTESKITATTQV 919
DB 901 TKTPAATKPEMTTAKOKTTERDLRTPTPTTAAAPKMTKATATTEKTESKITATTQV 960

QY 920 TSTTTQDTTPFKITTLTKTTLAPKVITTKITITTEIMNKEPEATAKPDATNSKATTPK 979
```

```
DB 961 TSTTTQDTTPFKITTLTKTTLAPKVITTKITITTEIMNKEPEATAKPDATNSKATTPK 1020
QY 980 PQKPTKAPKPTSTKPKTMTPRVRKPKXTTTPRKMSTMPELNPTSRIAEAMLTTRPN 1039
DB 1021 PQKPTKAPKPTSTKPKTMTPRVRKPKXTTTPRKMSTMPELNPTSRIAEAMLTTRPN 1080
QY 1040 QTFNSKLVEVNPXSSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLRVPRVNOGIIINPMLS 1099
DB 1081 QTFNSKLVEVNPXSSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLRVPRVNOGIIINPMLS 1140

RESULT 6
AAU32262
ID AAU32262 standard; protein; 1415 AA.
XX
AC AAU32262;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #2753.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
WU WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US008656.
XX
PR 18-APR-2000; 2000US-00552929.
PR 26-JAN-2001; 2001US-00770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
XX
PS Claim 20; Page 573; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU33304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
SQ Sequence 1415 AA;

Query Match
Best Local Similarity 99.8%; Score 5820.9; DB 4; Length 1415;
Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVLIQVSSQ----- 25
```

```
Db      12 MAWKTLPIYLILLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDF 71
Qy      26 -----ELSKGRCFESFERGECDDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 79
Db      72 KVCVTAELSCKRCFESFERGECDDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 131
Qy      80 PPSGASQTIKSTTKSPKPNKKTKVIESBEITEHVSVENQBSSESSSSSSSSSSSTI 139
Db      132 PPSGASQTIKSTTKSPKPNKKTKVIESBEITEHVSVENQBSSESSSSSSSSSTI 191
Qy      140 KIKSSKNSAANRELOKLVKDKNKNRTKKFTPKPPVVDVDEAGSLDNGDFKVTPTDST 199
Db      192 KIKSSKNSAANRELOKLVKDKNKNRTKKFTPKPPVVDVDEAGSLDNGDFKVTPTDST 251
Qy      200 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSILTVNKETVTKETTTINKQTSIDG 259
Db      252 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSILTVNKETVTKETTTINKQTSIDG 311
Qy      260 KEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTTPKEPAS 319
Db      312 KEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTTPKEPAS 371
Qy      320 TTPKEPTTTPKESAPTTKKEPAPTTTKSAPTTKKEPAPTTTKKEPAPTTTKKEP 379
Db      372 TTPKEPTTTPKESAPTTKKEPAPTTTKSAPTTKKEPAPTTTKKEPAPTTTKKEP 431
Qy      380 APTTKSAPTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTK 439
Db      432 APTTKSAPTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTK 491
Qy      440 EPAPTAPKAPPTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAP 499
Db      492 EPAPTAPKAPPTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAP 551
Qy      500 TTKSAPTTTKESPTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAP 559
Db      552 TTKSAPTTTKESPTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAP 611
Qy      560 APTAPKEPAPTTKETAAPTKPLTPTPEKLAPTTPKEPAPTTPEELAPTTPEEPTPT 619
Db      612 APTAPKEPAPTTKETAAPTKPLTPTPEKLAPTTPKEPAPTTPEELAPTTPEEPTPT 671
Qy      620 PEEPAPTTKKAAAPNTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKGAPT 679
Db      672 PEEPAPTTKKAAAPNTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKGAPT 731
Qy      680 APTTKKAPKELAPTTTKESPTTSKAPPTPKGTAPTTPKEPAPTTTKKEPAPTTKPG 739
Db      732 APTTKKAPKELAPTTTKESPTTSKAPPTPKGTAPTTPKEPAPTTTKKEPAPTTKPG 791
Qy      740 TAPTTLKEPAPTTKPKAPKELAPTTTKGPTSTTSKAPPTPKETAAPTTPKEPAPTTPK 799
Db      792 TAPTTLKEPAPTTKPKAPKELAPTTTKGPTSTTSKAPPTPKETAAPTTPKEPAPTTPK 851
Qy      800 KPAPTTPETPPPTSEVSTPTTKETPTTIHKSDESTPELSAPPTKALENSPKBPGVPT 859
Db      852 KPAPTTPETPPPTSEVSTPTTKETPTTIHKSDESTPELSAPPTKALENSPKBPGVPT 911
Qy      860 TKTPAAKPEMTTAKDKTTERDLRTTPEPTTAAKPVTKETAPTTEKTTESKITATTQV 919
Db      912 TKTPAAKPEMTTAKDKTTERDLRTTPEPTTAAKPVTKETAPTTEKTTESKITATTQV 971
Qy      920 TSITTTQDTPFKITLTKTTLAPKVTTKKTIITTEIMNKPETAAPKORATNSKATTPK 979
Db      972 TSITTTQDTPFKITLTKTTLAPKVTTKKTIITTEIMNKPETAAPKORATNSKATTPK 1031
Qy      980 PQKPTKAPKPTSTKPKTNPRVRKPTTTPPKVSTWPELNTSRIAEMLQTTTRN 1039
Db      1032 PQKPTKAPKPTSTKPKTNPRVRKPTTTPPKVSTWPELNTSRIAEMLQTTTRN 1091
Qy      1040 QTPNSKLVEVNPKSEDAAGGEGTTPHMLLRPHVFMPEVTPDMDYLRVPNQGIINPMLS 1099
```

```
Db      1092 QTPNSKLVEVNPKSEDAAGGEGTTPHMLLRPHVFMPEVTPDMDYLRVPNQGIINPMLS 1151
RESULT 7
ADK65839
ID      ADK65839 standard; protein; 1404 AA.
XX
AC      ADK65839;
XX
DT      06-MAY-2004 (first entry)
XX
DE      Angiogenesis-differentially expressed protein #53.
XX
KW      cytosolic; cardiant; vasotropic; antiarteriosclerotic;
KW      angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
KW      gene expression; cancer; coronary artery disease; myocardial ischemia;
KW      coronary arteriosclerosis; forensic medicine.
XX
OS      Homo sapiens.
XX
PN      WO2003066831-A2.
XX
PD      14-AUG-2003.
XX
PF      07-FEB-2003; 2003WO-US003848.
XX
PR      07-FEB-2002; 2002US-00067482.
PR      10-JUN-2002; 2002US-00164595.
PR      16-AUG-2002; 2002US-0403649P.
PR      03-JAN-2003; 2003US-0437746P.
XX
(PA      (ORIG-) ORIGENE TECHNOLOGIES INC.
XX      Sun Z, Li X, Kovacs KF, Fan W, Jay G;
XX      WPI; 2003-731502/69.
XX
PT      Determining the angiogenic index of a tissue or cell sample using
PT      expression levels of differentially expressed genes, useful for
PT      diagnosing or treating cancer, coronary artery disease, myocardial
PT      ischemia and/or arteriosclerosis.
XX
PS      Disclosure; SEQ ID NO 78; 296pp; English.
XX
CC      The invention relates to a method of determining the angiogenic index of
CC      a tissue or cell sample comprising assessing, in a sample, the expression
CC      levels of one or more differentially-expressed gene from any of 34 DNA
CC      sequences, given in the specification, where the levels are indicative of
CC      the angiogenic index. The methods and compositions of the present
CC      invention are useful for diagnosing, preventing and/or treating cancer,
CC      coronary artery disease, myocardial ischemia or coronary
CC      arteriosclerosis. They can also be used in research, drug discovery and
CC      forensic medicine involving angiogenesis. This sequence corresponds to
CC      one of the differentially expressed proteins of the invention.
XX
SQ      Sequence 1404 AA;
Query Match      99.4%; Score 5797.9; DB 7; Length 1404;
Best Local Similarity 96.1%; Pred. No. 2.8e-156;
Matches 1096; Conservative 0; Mismatches 3; Indels 41; Gaps 1;
Qy      1 MAWKTLPIYLILLLSVFIQVSSQ----- 25
Db      1 MAWKTLPIYLILLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCDYNCQHYMECCPDF 60
Qy      26 -----ELSKGRCFESFERGECDDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 79
Db      61 KVCVTAELSCKRCFESFERGECDDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
Qy      80 PPSGASQTIKSTTKSPKPNKKTKVIESBEITEHVSVENQBSSESSSSSSSSSTI 139
Db      121 PPSGASQTIKSTTKSPKPNKKTKVIESBEITEHVSVENQBSSESSSSSSSTI 180
```


Db 258 PALTTFKBPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 317
Qy 361 KBPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 420
Db 318 KBPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 377
Qy 421 TTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 480
Db 378 TTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 437
Qy 481 PAPTTTKSAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 540
Db 438 PAPTTTKSAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 497
Qy 541 KEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSA 600
Db 498 KEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSA 557
Qy 601 PTPPELAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 660
Db 558 PTPPELAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 617
Qy 661 KETAPTTPKGTAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 720
Db 618 KETAPTTPKGTAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 677
Qy 721 PKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 780
Db 678 PKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 737
Qy 781 TPKEAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 840
Db 738 TPKEAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 797
Qy 841 ABPTPKALENSKPEGVPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 900
Db 798 ABPTPKALENSKPEGVPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 857
Qy 901 ATTEKTESKATATTTQVSTTTQVSTTTQVSTTTQVSTTTQVSTTTQVSTTTQVSTTTQ 960
Db 858 ATTEKTESKATATTTQVSTTTQVSTTTQVSTTTQVSTTTQVSTTTQVSTTTQVSTTTQ 917
Qy 961 EETAKPKORATNSKATTPKPKQPTKAPKPTSTKPKPTMPVRKPKPTTPPKRMTSTMP 1020
Db 918 EETAKPKORATNSKATTPKPKQPTKAPKPTSTKPKPTMPVRKPKPTTPPKRMTSTMP 977
Qy 1021 LNPTRIAEAMLOTTTRPNQTSKLVNPNKXSEDAGGAGETPHMLLRPHVFMPEVTPD 1080
Db 978 LNPTRIAEAMLOTTTRPNQTSKLVNPNKXSEDAGGAGETPHMLLRPHVFMPEVTPD 1037
Qy 1081 MDYLPVNPQGIINPMLS 1099
Db 1038 MDYLPVNPQGIINPMLS 1056

Key Peptide Location/Qualifiers
1. .29
/label= Signal peptide
/note= "Spans residues 1 to 18, 20, 21, 24, 27 or 29 according to identification method"

WO2004013292-A2.
12-FEB-2004.
30-JUL-2003; 2003WO-US024084.
02-AUG-2002; 2002US-0400810P.
19-SEP-2002; 2002US-0412197P.
04-OCT-2002; 2002US-0416004P.
08-NOV-2002; 2002US-0424862P.
(INCY-) INCYTE CORP.
Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP, Richardson TW, Emerling BM, Lindquist EA, Chawla NK, Ramkumar J, Lee SY;
WPI; 2004-157116/15.
N-PSDB; ADK67916.
New extracellular messengers and nucleic acids, useful for diagnosing, treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or autoimmune thyroiditis.
Claim 59; SEQ ID NO 4; 165pp; English.
The present sequence is that of novel human extracellular messenger (EXMES) incyte ID NO: 7513017CD1 polypeptide. The protein shows homology to human megakaryocyte stimulating factor. The invention provides EXMES polynucleotides and polypeptides, as well as expression vectors, host cells, antibodies, agonists and antagonists, and methods for diagnosing, treating or preventing disorders associated with aberrant expression of EXMES, especially autoimmune and inflammatory disorders, cell proliferative disorders and endocrine disorders, e.g. adult respiratory distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's disease, diabetes mellitus, myasthenia gravis, osteoarthritis, osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal, parasitic, protozoal or helminthic infections, cancers, autoimmune thyroiditis, cretinism, Plummer's disease or thyroid carcinoma. Embodiments also provide methods for using the purified EXMES and/or their encoding polynucleotides for facilitating the drug discovery process, including determining of efficacy, dosage, toxicity and pharmacology, and for investigating the pathogenesis of diseases and medical conditions.

Sequence 1270 AA;
Query Match 91.8%; Score 5353.7; DB 8; Length 1270;
Best Local Similarity 91.5%; Pred. No. 9.7e-144;
Matches 1006; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
Qy 1 MAWKTLPIYLLLSVFVIOQVSSQELSCKGRGCFESFERGECDCDAQCKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLSVFVIOQVSSQELSCKGRGCFESFERGECDCDAQCKYDKCCPDYE 60
Qy 61 SFCAEVNHTSPSPSKKAPPSPGASQTIKSTTKRSPKPPNKKTKKVIIESEETEEHSVS 120
Db 61 SFCAE----- 65
Qy 121 ENQSSSSSSSSSSSSSIWIKSSKNSAANRELQKLKVKDNKNRTRKKPTPKPPVUDE 180
Db 66 -----VKDNKNRTRKKPTPKPPVUDE 87
Qy 181 AGSGLNDGFKVTTPTDSTTQHNVKSTSPKITTAKPINPRPSLPNSDTSKETSLTVNKE 240
Db 88 AGSGLNDGFKVTTPTDSTTQHNVKSTSPKITTAKPINPRPSLPNSDTSKETSLTVNKE 147

QY 80 PPSGASQTIKSTTKRSPKPNKKTKKVI BSEETIEHSVSENQESSSSSSSSSTI W 139
 Db 107 ----- 106
 QY 140 KIKSSKNSAANRELQKLVKXNDKNKRTKKXPTKPPVWDEAGSLDNGDPKVTTPDTST 199
 Db 107 -----VKDNKNRTKKXPTKPPVWDEAGSLDNGDFKVTTPDTST 147
 QY 200 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLTVNKETTVETKETTNNKQSTG 259
 Db 148 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLTVNKETTVETKETTNNKQSTG 207
 QY 260 KEKTSASKEQSIKTSADLAPSKVLAKPTPKAETTTKGPALTTPKEPTPTPKBPAS 319
 Db 208 KEKTSASKEQSIKTSADLAPSKVLAKPTPKAETTTKGPALTTPKEPTPTPKBPAS 267
 QY 320 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 379
 Db 268 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 327
 QY 380 APTTTKSAPTTKPEAPTTTKKPAETTPKPEAPTTKPEPTTTKPEAPTTKPEAPTTPK 439
 Db 328 APTTTKSAPTTKPEAPTTTKKPAETTPKPEAPTTKPEPTTTKPEAPTTKPEAPTTPK 387
 QY 440 EPAPTAPKPAETTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAP 499
 Db 388 EPAPTAPKPAETTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAP 447
 QY 500 TTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKKPAETTPKPEAPTTKPEAPTTTKK 559
 Db 448 TTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKKPAETTPKPEAPTTTKPEAPTTTKK 507
 QY 560 APTAPEKAPTTKETAETTPKPLNPTTPEKLAPTTPEKPAETTPPEELAPTTPEEPTTT 619
 Db 508 APTAPEKAPTTKETAETTPKPLNPTTPEKLAPTTPEKPAETTPPEELAPTTPEEPTTT 567
 QY 620 PEEAPTTPKAAAPNTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKGTAPTTKPE 679
 Db 568 PEEAPTTPKAAAPNTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKGTAPTTKPE 627
 QY 680 APTPKKPAKELAPTTTKPEPTSTSDKPAETTPKGTAPTTKPEAPTTKPEAPTTKPKG 739
 Db 628 APTPKKPAKELAPTTTKPEPTSTSDKPAETTPKGTAPTTKPEAPTTKPEAPTTKPKG 687
 QY 740 TAPTTLKPEAPTTPKKPAKELAPTTTKGPTSTSDKPAETTPKETAETTPKPEAPTTPK 799
 Db 688 TAPTTLKPEAPTTPKKPAKELAPTTTKGPTSTSDKPAETTPKETAETTPKPEAPTTPK 747
 QY 800 KPAPTTPEPTTSEVSTPTTKPEPTTIHKSPDESTBELSAPTPKALENSPKPEGVT 859
 Db 748 KPAPTTPEPTTSEVSTPTTKPEPTTIHKSPDESTBELSAPTPKALENSPKPEGVT 807
 QY 860 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAAPKWTKETATTTTEKTTESKITATTQV 919
 Db 808 TKTPAAKPEMTTAKDKTTERDLRTTPETTTAAAPKWTKETATTTTEKTTESKITATTQV 867
 QY 920 TSTTTQDTTFFKITLTKTTLAPKVTITTKITTTIMNKPEETAKPKDRATNSKATTPK 979
 Db 868 TSTTTQDTTFFKITLTKTTLAPKVTITTKITTTIMNKPEETAKPKDRATNSKATTPK 927
 QY 980 PQKPTKAPKPTSTKPKTNPVRKKTTPTRKMTSTMPELNPTSRIAEAMLQTTTRPN 1039
 Db 928 PQKPTKAPKPTSTKPKTNPVRKKTTPTRKMTSTMPELNPTSRIAEAMLQTTTRPN 987
 QY 1040 QTPNSKLVEVNPKSDEAGGAGETPHMLLRPHVFMPEVTPDMDYLPVRVNOGIIINPMLS 1099
 Db 988 QTPNSKLVEVNPKSDEAGGAGETPHMLLRPHVFMPEVTPDMDYLPVRVNOGIIINPMLS 1047

RESULT 11
 ABUS3252
 ID ABUS3252 standard; protein; 546 AA.
 XX

ABUS3252;
 14-APR-2003 (first entry)
 Human testes-derived DKFZphtes3_4019 homologue #1.
 Human; gene therapy; vaccine; disease treatment; detection.
 Homo sapiens.
 WO2001112659-A2.
 22-FEB-2001.
 18-AUG-2000; 2000WO-IB001496.
 18-AUG-1999; 99US-0149499P.
 28-SEP-1999; 99US-0156503P.
 (GEHU-) GERMAN HUMAN GENOME PROJECT.
 Wiemann S;
 WPI; 2001-327840/34.
 Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies.
 Example III; Page 892; 1095pp; English.
 This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence represents a homologue of a polypeptide described in the disclosure of the invention

Sequence 546 AA;
 Query Match 50.0%; Score 2920; DB 4; Length 546;
 Best Local Similarity 100.0%; Pred. No. 3.5e-75;
 Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 447 KKPAETTPKPAETTPKPEAPTTTKPEPTTTKPEAPTTTKSAPTTTKPEAPTTTKSAPT 506
 Db 1 KKPAETTPKPAETTPKPEAPTTTKPEPTTTKPEAPTTTKSAPTTTKPEAPTTTKSAPT 60
 QY 507 TPKEPSPTTKPEAPTTKPEAPTTPKKPAETTPKPEAPTTKPEAPTTTKKPAETPK 566
 Db 61 TPKEPSPTTKPEAPTTKPEAPTTPKKPAETTPKPEAPTTKPEAPTTTKKPAETPK 120
 QY 567 PAPTTPKETAETTPKELAPTTPEKLAPTTPEKAPTTPEELAPTTPEEPTTTPEEAPT 626
 Db 121 PAPTTPKETAETTPKELAPTTPEKLAPTTPEKAPTTPEELAPTTPEEPTTTPEEAPT 180
 QY 627 TPKAAAPNTPKPEAPTTKPEAPTTKPEAPTTPKETAETTPKETAETTPKETAETTPK 686
 Db 181 TPKAAAPNTPKPEAPTTKPEAPTTKPEAPTTPKETAETTPKETAETTPKETAETTPK 240
 QY 687 PAPKELAPTTTKETSSTSDKPAETTPKGTAPTTKPEAPTTKPEAPTTKGTAPTTK 746
 Db 241 PAPKELAPTTTKETSSTSDKPAETTPKGTAPTTKPEAPTTKPEAPTTKGTAPTTK 300
 QY 747 EPAPTTPKKPAKELAPTTTKGPTSTSDKPAETTPKETAETTPKETAETTPKETAETTP 806
 Db 301 EPAPTTPKKPAKELAPTTTKGPTSTSDKPAETTPKETAETTPKETAETTPKETAETTP 360
 QY 807 ETPPPTTSEVSTPTTKPEPTTIHKSPDESTBELSAPTPKALENSPKPEGVTTKTTPAAT 866

```

Db 361 ETPTPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAETPKALENSPKPEPGVPTTKTPAAT 420
QY 867 KPEMTTAKDKTERDLRTTPTTTAAAPKMTKETATTTTEKTESKITATTTQVSTTTQD 926
Db 421 KPEMTTAKDKTERDLRTTPTTTAAAPKMTKETATTTTEKTESKITATTTQVSTTTQD 480
QY 927 TTPFKITLTKTTLAPKVTTTKTITTTTEIMNKPEETAKPKDRATNSKATTPKPKPTKA 986
Db 481 TTPFKITLTKTTLAPKVTTTKTITTTTEIMNKPEETAKPKDRATNSKATTPKPKPTKA 540
QY 987 PKKPTS 992
Db 541 PKKPTS 546

RESULT 12
ABUS3253
ID ABUS3253 standard; protein; 551 AA.
XX AC ABUS3253;
XX DT 14-APR-2003 (first entry)
XX DE Human testes-derived DKFZphtes3_4c19 homologue #2.
XX KW Human; gene therapy; vaccine; disease treatment; detection.
XX OS Homo sapiens.
XX PN WO200112659-A2.
XX PD 22-FEB-2001.
XX PF 18-AUG-2000; 2000WO-IB001496.
XX PR 18-AUG-1999; 99US-0149499P.
XX PR 28-SEP-1999; 99US-0156503P.
XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX PI Wiemann S;
XX DR WPI; 2001-327840/34.
XX PT Nucleic acids having the sequences of clones isolated from libraries of
XX different human tissues, useful in recombinant DNA methodologies.
XX PS Example III; Page 892-893; 1095pp; English.
XX CC This invention describes novel polynucleotides and polypeptides isolated
CC from human cDNA libraries which can be used for gene therapy or in
CC vaccines. The polynucleotides of the invention and antibodies encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. The products of the
CC invention may also be used to identify modulators of expression and
CC activity and to down regulate expression and activity. The antibodies of
CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence represents a homologue
CC of a polypeptide described in the disclosure of the invention
XX SQ Sequence 551 AA;

Query Match 49.5%; Score 2888; DB 4; Length 551;
Best Local Similarity 100.0%; Pred. No. 2.9e-74;
Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 556 TKKPAPAPKEPAPTTPKETAPTPKLTPTTPKLAPTTPEKPAPTTPELAPTTPEEP 615
Db 1 TKKPAPAPKEPAPTTPKETAPTPKLTPTTPKLAPTTPEKPAPTTPELAPTTPEEP 60

QY 616 TPTTPEAPPTPKAAAPNTPKSPAPTTPKPEAPTTPKPEAPTTPKETAPTTPKGTAPTT 675
Db 61 TPTTPEAPPTPKAAAPNTPKSPAPTTPKPEAPTTPKPEAPTTPKETAPTTPKGTAPTT 120

```

```

QY 676 LKEPAPTTPKKPAKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTKPEAPTTKPEAPT 735
Db 121 LKEPAPTTPKKPAKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTKPEAPTTKPEAPT 180
QY 736 TPKGTAPTTLKEPAPTTPKKPAKELAPTTTKEPTSTTSDKPAPTTKPEAPTTPKPEAP 795
Db 181 TPKGTAPTTLKEPAPTTPKKPAKELAPTTTKEPTSTTSDKPAPTTKPEAPTTPKPEAP 240
QY 796 TTPKKAPAPTTPEPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALENSPKPEP 855
Db 241 TTPKKAPAPTTPEPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALENSPKPEP 300
QY 856 GVPTTTPAATKPEMTTAKDKTERDLRTTPTTTAAAPKMTKETATTTTEKTESKITAT 915
Db 301 GVPTTTPAATKPEMTTAKDKTERDLRTTPTTTAAAPKMTKETATTTTEKTESKITAT 360
QY 916 TTVSTSTTTQDTPFKITLTKTTLAPKVTTTKTITTTTEIMNKPEETAKPKDRATNSKA 975
Db 361 TTVSTSTTTQDTPFKITLTKTTLAPKVTTTKTITTTTEIMNKPEETAKPKDRATNSKA 420
QY 976 TTPKPKPTKAPKPTSTTKPKTMTPEVRKPKTTPRKMSTMPBLNPTSRIAEAMLOTT 1035
Db 421 TTPKPKPTKAPKPTSTTKPKTMTPEVRKPKTTPRKMSTMPBLNPTSRIAEAMLOTT 480
QY 1036 TRPNQTPNSKLVEVNPKSDAGAGETPHMLLRPHVFMPEVTPDMVLPVFNQGIIN 1095
Db 481 TRPNQTPNSKLVEVNPKSDAGAGETPHMLLRPHVFMPEVTPDMVLPVFNQGIIN 540
QY 1096 PMLS 1099
Db 541 PMLS 544

RESULT 13
AAB29778
ID AAB29778 standard; protein; 902 AA.
XX AC AAB29778;
XX DT 28-FEB-2001 (first entry)
XX DE Human MSF-derived triboneectin.
XX KW Human triboneectin; MSF; megakaryocyte stimulating factor;
XX alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
XX osteoarthritis; tribosupplementation; tissue adhesion inhibition;
XX friction coefficient reduction; gene therapy; antiarthritic; osteopathic.
XX OS Homo sapiens.
XX PN WO200064930-A2.
XX PD 02-NOV-2000.
XX PF 24-APR-2000; 2000WO-US010953.
XX PR 23-APR-1999; 99US-00298970.
XX PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
XX PI Jay GD;
XX DR WPI; 2001-024673/03.
XX PT Novel triboneectin polypeptide useful as lubricant for treating
XX osteoarthritis, comprises O-linked lubricating moiety.
XX PS Disclosure; Fig 1; 47pp; English.
XX CC The invention relates to a human triboneectin which is a product of
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
CC gene. The triboneectin has at least one O-linked oligosaccharide

```


CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats
CC of a motif having at least 50% identity to the sequence KEPAPTT
CC (AA29774). The invention also relates to a nucleic acid encoding a human
CC MSF-derived tribonection; a bio-compatible composition comprising a human
CC tribonection for inhibiting tissue adhesion formation; and a method of
CC diagnosing osteoarthritis or a predisposition to osteoarthritis by
CC measuring the amount of MSF or its fragment in a biological sample of a
CC mammal, wherein an increased amount of MSF compared to a control
CC indicates the presence of or predisposition to developing osteoarthritis.
CC The tribonection and DNA encoding it are useful in the treatment of
CC osteoarthritis, where they may be used for lubricating mammalian joints,
CC such as articulating joints of humans, dogs or horses. The tribonection,
CC when formulated as a membrane, foam, gel or fibre, is useful for
CC inhibiting adhesion between two surfaces such as the injured tissues of a
CC mammal where the injury is caused by a surgical insertion or trauma, or
CC an artificial device e.g., an orthopaedic implant. In particular, one of
CC the surfaces is pericardial tissue. DNA encoding a tribonection may be
CC used in gene therapy. The present sequence represents a substantial
CC portion of a human MSF-derived tribonection
XX
XX
SQ Sequence 902 AA;

Query Match 49.4%; Score 2880.1; DB 4; Length 902;
Best Local Similarity 71.4%; Pred. No. 9.3e-74;
Matches 675; Conservative 27; Mismatches 74; Indels 169; Gaps 66;

QY	159	VKDNNKQRTKKKTPKPPVVDVAGSLDNGDFKVTTPDTSTTQHNKSVTSKITTAKPIN	218
Db	1	VKDNNKQRTKKKTPKPPVVDVAGSLDNGDFKVTTPDTSTTQHNKSVTSKITTAKPIN	60
QY	219	PPPSLPNSDTSKETSLSLVNKTETTTNNKQTSDDGKEKTSKETSOSIEKTSK	278
Db	61	PPPSLPNSDTSKETSLSLVNKTETTTNNKQTSDDGKEKTSKETSOSIEKTSK	120
QY	279	DIAPTSKVLAQTPKPAETTKGPAETTKPEPTTPKPEASTTPKEPTPTTIKSAPTTPK	338
Db	121	DIAPTSKVLAQTPKPAETTKGPAETTKPEPTTPKPEASTTPKEPTPTTIKSAPTTPK	172
QY	339	EPAPTTKSAPTTKPEAPTTTKEAPTTKPEAPTTTKEAPTTTKEAPTTTKEAPTTT	398
Db	173	EPAPTTKSAPTTKPEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTT	228
QY	399	PKKPAETTKPEAPTTPKPPTTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEPA	458
Db	229	-KEPAPTTT-KEPAPTTT-KEPAPTTT-KEPAPTTT-KEPAPTTT-KEPAPTTT-KEPA	281
QY	459	PTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE	518
Db	282	PIT-KEPAP-TTKEPAPTTT-KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE	335
QY	519	PAPTTPKEPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT	578
Db	336	PAPTTP-KEPAPTTT-KEPAPTTT-KEPAPTTT-KEPAPTTT-KEPAPTTT-KEPAPT	388
QY	579	TPKLLPTTPEKLAPTTPEKPAETTPPELAPTTPEEPPTTTPPEAPTTTPEKAAAPTKE	638
Db	389	T-KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE	440
QY	639	PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	596
Db	441	PAPTTP-KEPAPTTT-KEPAPTTT-KEPAPTTT-KEPAPTTT-KEPAPTTT-KEPAPT	493
QY	697	TKPSTSTSDKPAETTKGAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE	756
Db	494	TKPAPTTKE-PAPTTP-KEPAPTTT-KEPAPTTT-KEPAPTTT-KEPAPTTT-KEP	546
QY	757	AP--KEAPTTTKEGPTSTSDKPAETTKGAPTTTKEPAPTTTKEPAPTTTKEPAPT	814
Db	547	APTTPKEPAP-TTKEPAPTTKE-PAPTTP-KEPAPTTT-KEPAPTTT-KEPAPTTTKE	600
QY	815	EVSTPTTTKEPTTHKSPDESTELSAPETPKALENSPKPGVPTTKTPAATKPEMTT	874
Db	601	EPA--PTTKGPAETTKGAPTTT-----KEPAP-----TTKEP-APTTPKEPAPT	640

QY	875	KOKTTERDLRTTPTTTTAAAPKMTKETATTEKTSKITATTQTQVSTTTQDTTPPKITT	934
Db	641	-----TKEPAPTT-----KEPAPTTKEPAPTTKEPAP-----	667
QY	935	LKTTTLAPKVTTTKTITTEIMNKPEETAKPKDRATNSKATTEPKPKTKAPKPKTSTK	994
Db	668	-----TTKEPAPTTKEP-APT-TK	684
QY	995	KPKTMRVRKPKTTPTRKWTSTMPBLNPTSRIAEAMLQTTTRPNQTPNSKLVENVNPKGE	1054
Db	685	EP-----APTTPRKTWTSTMPBLNPTSRIAEAML-TTTRPNQTPNSKLVENVNPKGE	735
QY	1055	DAGGAGETPHMLLRPHVMPVETPPDMDYLPRVNOGIIINPMLS	1099
Db	736	DAGGAGETPHMLLRPHVMPVETPPDMDYLPRVNOGIIINPMLS	780

RESULT 14

ABUS3254
ID ABUS3254 standard; protein; 513 AA.

XX	AC	ABUS3254;
XX	AC	
DT	14-APR-2003	(first entry)
XX	Human testes-derived DKFzhtes3_4019	homologue #3.
XX	Human; gene therapy; vaccine; disease treatment; detection.	
XX	Homo sapiens.	
XX	WO200112659-A2.	
XX	22-FEB-2001.	
XX	18-AUG-2000; 2000WO-IB001496.	
XX	18-AUG-1999; 99US-0149499P.	
XX	28-SEP-1999; 99US-0156503P.	
XX	(GEHU-) GERMAN HUMAN GENOME PROJECT.	
XX	Wiemann S;	
XX	WPI; 2001-327840/34.	

Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies.

Example III; Page 893; 1095pp; English.

This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence represents a homologue of a polypeptide described in the disclosure of the invention

Sequence 513 AA;

Query Match 47.2%; Score 2757; DB 4; Length 513;

Best Local Similarity 100.0%; Pred. No. 1.4e-70;

Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	166	RTKKKPTKPPVVDVAGSLDNGDFKVTTPDTSTTQHNKSVTSKITTAKPINRPSLPP	225
Db	1	RTKKKPTKPPVVDVAGSLDNGDFKVTTPDTSTTQHNKSVTSKITTAKPINRPSLPP	60

QY 226 NSDTSKETSLSLVNKEVETVETKETTNNKQSTGKCKTTSKETSQIENKSAKDLAPTSPK 285
 DB 61 NSDTSKETSLSLVNKEVETVETKETTNNKQSTGKCKTTSKETSQIENKSAKDLAPTSPK 120
 QY 286 VLAKPTPKAETTTKGPALTTKPEPTTTPKEPASTTPKEPTTTTTSKAPTTPKEPAPTTT 345
 DB 121 VLAKPTPKAETTTKGPALTTKPEPTTTPKEPASTTPKEPTTTTTSKAPTTPKEPAPTTT 180
 QY 346 KSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 405
 DB 181 KSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 240
 QY 406 TPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 465
 DB 241 TPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 300
 QY 466 APTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 525
 DB 301 APTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 360
 QY 526 EPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 585
 DB 361 EPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 420
 QY 586 TTPKELAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 645
 DB 421 TTPKELAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 480
 QY 646 EPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 678
 DB 481 EPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 513

RESULT 15

AA080041
 ID AAR80041 standard; protein; 452 AA.

XX AC AAR80041;

XX DT 25-MAR-2003 (revised)

XX DT 10-APR-1996 (first entry)

XX DE Human megakaryocytopoietin protein.

XX KW Human; megakaryocytopoietin; wheat germ agglutinin; heparin;
 KW megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;
 KW multipotential stem cell.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 FT Misc-difference 393..396
 FT /note= "unspecified amino acids"

FT FT Misc-difference 444..446
 FT /note= "unspecified amino acids"

XX W09523861-A1.
 XX 08-SEP-1995.

XX 06-MAR-1995; 95WO-CN000015.

XX 04-MAR-1994; 94CN-00112066.

XX (SHAN-) SHANGHAI BEITE BIOTECHNOLOGY CO LTD.
 XX Gu X, Han Z, Shen Q;

XX WPI; 1995-320576/41.
 XX N-PSD5; AAT04546.

XX PT New haematopoietic cell growth factor - used for treating
 PT thrombocytopenia and hematocytopenia.

XX Example; Page 23; 36pp; Chinese.

XX This sequence represents the human megakaryocytopoietin (MPO) protein.
 CC This sequence was purified using a carrier which can couple wheat germ
 CC agglutinin and heparin to separate MPO. Fragments of this sequence (see
 CC AAR80039 and AAR80040) were used to produce the amplification primers
 CC shown in AAT04544 and AAT04545. The fragments amplified by these primers
 CC can then be used as probes to screen human cDNA libraries for MPO cDNA.
 CC The MPO cDNA can then be inserted into a plasmid which is used to
 CC transform cells to produce MPO. The MPO sequence is capable of promoting
 CC colony formation of megakaryocytes, enlarging the size of megakaryocytes
 CC and stimulating the proliferation of multipotential stem cells. The
 CC factor may be used for treating thrombocytopenia and hematocytopenia. The
 CC purification method can be used to isolate MPO from human urine or serum
 CC of patients with aplastic anaemia, and from animal blood or urine by
 CC radiation exposing the animals to induce aplastic anaemia. (Updated on 25
 CC -MAR-2003 to correct PA field.)

XX SQ Sequence 452 AA;

Query Match 33.5%; Score 1954.5; DB 2; Length 452;

Best Local Similarity 76.1%; Pred. No. 6.8e-48;

Matches 378; Conservative 3; Mismatches 21; Indels 95; Gaps 3;

QY 1 MAWKTLPIVLLLLSVFVIQVSSQ----- 25

DB 1 MAWKTLPIVLLLLSVFVIQVSSQDLSSCAGRCGSGYSRDATCNDYNCQHYMECCPDF 60

QY 26 -----ELSCKGRCPESFERGECDCDAQCKYDKCCPDYSEFCAEVHNPTSPSSSKAP 79

DB 61 KRVTAEELSCKGRCPESFERGECDCDAQCKYDKCCPDYSEFCAEVHNPTSPSSSKAP 120

QY 80 PPSGASQTIKSTTKGSPKPPNKKTKVIESBEITEHSVSENQSSSSSSSSSSSTIW 139

DB 121 PPSGASQTIKSTTKGSPKPPNKKTKVIESBEITE----- 156

QY 140 KTKSSKNANRELOKKLVKONKQRTKKKTPKPPVVDGAGSLDNGDFKVTPTDST 199

DB 157 -----VKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTPTDST 197

QY 200 TQHNKVSTSPKITTAKPINPRSLPNSDTSKETSLSLVNKEVETVETKETTNNKQSTG 259

DB 198 TQHNKVSTSPKITTAKPINPRSLPNSDTSKETSLSLVNKEVETVETKETTNNKQSTG 257

QY 260 KEKTTSAKETQSIENKSAKDLAPTSPKLAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 319

DB 258 KEKTTSAKETQSIENKSAKDLAPTSPKLAETTKGPAETTKGPAETTKGPAETTKGPAETTKGPA 317

QY 320 TTPKEPTPTTIKSAPTTPKEPAPTTTPKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 379

DB 318 TTPKEPTPTTIKSAPTTPKEPAPTTTPKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 377

QY 380 APTTTKSAPTTTPKEPAPTTTPKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTPK 439

DB 378 APTTTKSHPPPLPRSCXXXCTQP-----TPKEPHPLPRSLHPPKKEPAPTTPK 426

QY 440 EPAPTAPKPKAPTTTKE 456

DB 427 EPAPTAPKPKAPLPLE 443

Search completed: October 13, 2004, 11:36:40
 Job time : 96.6045 secs

cl1 (no alignment)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 19.1039 Seconds
(without alignments)
3615.116 Million cell updates/sec

Title: SEQ1-A
Perfect score: 5835
Sequence: 1 NAWKTLPIYLLLLSVFVIQ.....DMDYLRVNPNGIINPMLS 1099

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/prodata/1/1aa/5A-COMB.pep: *
2: /cgn2_6/prodata/1/1aa/5B-COMB.pep: *
3: /cgn2_6/prodata/1/1aa/5A-COMB.pep: *
4: /cgn2_6/prodata/1/1aa/5B-COMB.pep: *
5: /cgn2_6/prodata/1/1aa/PCTUS-COMB.pep: *
6: /cgn2_6/prodata/1/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5835	100.0	1363	4	US-07-757-022B-52
2	5820.9	99.8	1140	4	US-07-757-022B-104
3	5820.9	99.8	1404	4	US-07-757-022B-2
4	5820.9	99.8	1404	4	US-07-757-022B-62
5	5820.9	99.8	1404	4	US-09-298-970A-1
6	5797.9	99.4	1404	4	US-10-164-595-78
7	5617.7	96.3	1320	4	US-07-757-022B-46
8	5617.7	96.3	1320	4	US-07-757-022B-60
9	5608.7	96.1	1320	4	US-10-164-595-58
10	5603.6	96.0	1361	4	US-07-757-022B-40
11	5561	95.3	1049	4	US-07-757-022B-58
12	5561	95.3	1313	4	US-07-757-022B-142
13	5546.9	95.1	1354	4	US-07-757-022B-48
14	5377.9	93.2	1314	4	US-07-757-022B-50
15	5353.7	91.8	1038	4	US-07-757-022B-74
16	5353.7	91.8	1270	4	US-07-757-022B-44
17	5339.6	91.5	1311	4	US-07-757-022B-84
18	5331.9	89.7	1022	4	US-07-757-022B-14
19	5011	85.9	941	4	US-09-538-092-1258
20	1354.4	23.2	5179	4	US-07-757-022B-116
21	1331.8	19.4	891	4	US-08-714-741-32
22	923.9	15.8	220	4	US-07-757-022B-96
23	864.6	14.8	207	4	US-07-757-022B-136
24	864.6	14.8	207	4	US-07-757-022B-136
25	862.2	14.8	422	4	US-07-757-022B-94
26	861.9	14.8	209	4	US-07-757-022B-68
27	856.9	14.7	208	4	US-07-757-022B-132

ALIGNMENTS

RESULT 1
US-07-757-022B-52
; Sequence 52, Application US/07757022B
; Patent No. 5433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Gesner, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1363 amino acids
; TYPE: AMINO ACID

28	855.7	14.7	3256	4	US-09-919-172-98	Sequence 98, Appl
29	855.7	14.7	3256	4	US-09-976-594-22	Sequence 22, Appl
30	855.7	14.7	3256	4	US-09-919-039-21	Sequence 21, Appl
31	848.1	14.5	463	4	US-07-757-022B-54	Sequence 54, Appl
32	845.7	14.5	2972	3	US-09-579-181-2	Sequence 2, Appl
33	845.7	14.5	3118	3	US-09-579-181-1	Sequence 1, Appl
34	835.9	14.3	204	4	US-07-757-022B-92	Sequence 92, Appl
35	823.1	14.1	236	4	US-07-757-022B-70	Sequence 70, Appl
36	811.5	13.9	231	4	US-07-757-022B-30	Sequence 30, Appl
37	792.1	13.6	2142	4	US-09-538-092-1142	Sequence 1142, Ap
38	785.9	13.5	1837	3	US-08-928-361B-5	Sequence 5, Appl
39	785.9	13.5	1837	4	US-09-588-995A-5	Sequence 90, Appl
40	776.9	13.3	192	4	US-07-757-022B-90	Sequence 5, Appl
41	744.4	12.8	1721	3	US-08-700-651-5	Sequence 6, Appl
42	743.4	12.7	1721	3	US-08-928-361B-6	Sequence 6, Appl
43	743.4	12.7	1721	4	US-09-588-995A-6	Sequence 66, Appl
44	734.1	12.6	423	4	US-07-757-022B-66	Sequence 5, Appl
45	727.9	12.5	3969	3	US-08-061-376-5	

TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-52

Query Match 100.0%; Score 5835; DB 4; Length 1363;
Best Local Similarity 100.0%; Pred. No. 4.5e-170;
Matches 1099; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSCGRCFESFERGRCDCDAOCKYDKCCPDYE 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQELSCGRCFESFERGRCDCDAOCKYDKCCPDYE 60
QY 61 SCAEVHNTSPSSKAPPPSGASQTIKSTTKRSGKPPNKKTKKVIKESBEITEHSVS 120
DB 61 SCAEVHNTSPSSKAPPPSGASQTIKSTTKRSGKPPNKKTKKVIKESBEITEHSVS 120
QY 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKDKNKNRTKKXPTPKPPVDE 180
DB 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKDKNKNRTKKXPTPKPPVDE 180
QY 181 AGSLONGKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKE 240
DB 181 AGSLONGKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKE 240
QY 241 TTVEKETITTKNTDQKETSQKETSQKETSQKETSQKETSQKETSQKETSQKETSQK 300
DB 241 TTVEKETITTKNTDQKETSQKETSQKETSQKETSQKETSQKETSQKETSQKETSQK 300
QY 301 PALTTKEPTTTPKEPASTTTPKEPTTTPKEPASTTTPKEPASTTTPKEPASTTTP 360
DB 301 PALTTKEPTTTPKEPASTTTPKEPTTTPKEPASTTTPKEPASTTTPKEPASTTTP 360
QY 361 KEPAATTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKE 420
DB 361 KEPAATTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKE 420
QY 421 TTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTP 480
DB 421 TTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTP 480
QY 481 PAPTTKSAPTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTP 540
DB 481 PAPTTKSAPTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTP 540
QY 541 KEPAATTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKE 600
DB 541 KEPAATTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKE 600
QY 601 PTPPELAPTTPEPTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKE 660
DB 601 PTPPELAPTTPEPTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKE 660
QY 661 KETAPTTKGTAPTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKE 720
DB 661 KETAPTTKGTAPTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKE 720
QY 721 PKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKE 780
DB 721 PKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKE 780
QY 781 TPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKE 840
DB 781 TPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKE 840
QY 841 AEPTPKALENSPKPGVPTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKE 900
DB 841 AEPTPKALENSPKPGVPTTTPKEPASTTTPKEPASTTTPKEPASTTTPKEPASTTTPKE 900
QY 901 ATTTEKTESKITATTQVSTTQDTPPKITLKTLLAPKVTTPKKTITTEIMNKP 960
DB 901 ATTTEKTESKITATTQVSTTQDTPPKITLKTLLAPKVTTPKKTITTEIMNKP 960
QY 961 EETAKPKDRATNSKATTPKPKPTKPKPKPTKPKPKPTKPKPKPTKPKPKPTTTPPKWTSTMP 1020
```

```
DB 961 EETAKPKDRATNSKATTPKPKPTKPKPKPTKPKPKPTKPKPKPTTTPPKWTSTMP 1020
QY 1021 LNTSRIAEAMLOTTTRPNOTPNSKLVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 1080
DB 1021 LNTSRIAEAMLOTTTRPNOTPNSKLVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 1080
QY 1081 MDYLPRVNOGIIINPMLS 1099
DB 1081 MDYLPRVNOGIIINPMLS 1099

RESULT 2
US-07-757-022B-104
; Sequence 104, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseriz, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO. 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-104

Query Match 99.8%; Score 5820.9; DB 4; Length 1140;
Best Local Similarity 96.4%; Pred. No. 9.7e-170;
Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSCGRCGEGYSRDATCNDYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQELSCGRCGEGYSRDATCNDYNCQHYMECCPDF 60
```

Qy	26	-----ELSCKGRCFSFRGRGECDDACKKYDKCCPDYSEFCABVHNPTPPSSKKAP	79
Db	61	KRYCTAELSGKGRCSFRGRGECDDAQCKYDKCCPDYSEFCABVHNPTPPSSKKAP	120
Qy	80	PPSGASOTTIKSTTKRSPKPNKKTKKVIJESBEITEHSHSVSNQSSSSSSSSSSSSSTIW	139
Db	121	PPSGASOTTIKSTTKRSPKPNKKTKKVIJESBEITEHSHSVSNQSSSSSSSSSSSSSTIW	180
Qy	140	KIKSSKNSAANRELQKGLKVKONKXORTKKKPTPKPPVVDEAGSLDNGDFKVTTPTDTS	199
Db	181	KIKSSKNSAANRELQKGLKVKONKXORTKKKPTPKPPVVDEAGSLDNGDFKVTTPTDTS	240
Qy	200	TQHNKYVSTSPKITTAKPINPRPSLPNSDTSKETSUTVNKEITVETKETTNTTKNOTSTDG	259
Db	241	TQHNKYVSTSPKITTAKPINPRPSLPNSDTSKETSUTVNKEITVETKETTNTTKNOTSIDG	300
Qy	260	KEKTTSAKETQSLEKTSADLAPTCKVLAKPTPKAETTTKGPAITTPKEPTTTTKEPAS	319
Db	301	KEKTTSAKETQSLEKTSADLAPTCKVLAKPTPKAETTTKGPAITTPKEPTTTTKEPAS	360
Qy	320	TTPKESPTPTTIKSAPTTPKEPAPTITTKSAPTTPKEPAPTITTKSAPTTPKEPAPTITTKP	379
Db	361	TTPKESPTPTTIKSAPTTPKEPAPTITTKSAPTTPKEPAPTITTKSAPTTPKEPAPTITTKP	420
Qy	380	APTITTKSAPTTPKEPAPTTPKKAPITTPKEPAPTTPKEPTTPPKEPAPTTPKEPAPTTPK	439
Db	421	APTITTKSAPTTPKEPAPTTPKKAPITTPKEPAPTTPKEPTTPPKEPAPTTPKEPAPTTPK	480
Qy	440	EPAPTAKKPAPTTPKEPAPTTPKEPAPTTPTKKESPTTPPKEPAPTTPKSAPTTPKEPAPT	499
Db	481	EPAPTAKKPAPTTPKEPAPTTPKEPAPTTPTKKESPTTPPKEPAPTTPKSAPTTPKEPAPT	540
Qy	500	TTKSAPTTPKESPTPTTKAPAPTTPKEPAPTTPKKAPITTPKEPAPTTPKEPAPTTPKAP	559
Db	541	TTKSAPTTPKESPTPTTKAPAPTTPKEPAPTTPKKAPITTPKEPAPTTPKEPAPTTPKAP	600
Qy	560	APTAPKEPAPTTPKETAPTTPKKLTPTPKLAPTTPKEPAPTTPBELAPTTPPEPTPTT	619
Db	601	APTAPKEPAPTTPKETAPTTPKKLTPTPKLAPTTPKEPAPTTPBELAPTTPPEPTPTT	660
Qy	620	PEBPAPTTPKAAAPNTPKAPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEP	679
Db	661	PEBPAPTTPKAAAPNTPKAPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEP	720
Qy	680	APTTPKPAKELAPTTPKESPTSTSDKAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGT	739
Db	721	APTTPKPAKELAPTTPKESPTSTSDKAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGT	780
Qy	740	TAPTTLKEPAPTTPKKAPKELAPTTPKGTSTSDKAPTTPKGTAPTTPKEPAPTTPKAP	799
Db	781	TAPTTLKEPAPTTPKKAPKELAPTTPKGTSTSDKAPTTPKGTAPTTPKEPAPTTPKAP	840
Qy	800	KPAPTTPETPPPTTSEVSTPTTTKEPTTIHKSPPDESPELSAAPTPKALENSPKPGVPT	859
Db	841	KPAPTTPETPPPTTSEVSTPTTTKEPTTIHKSPPDESPELSAAPTPKALENSPKPGVPT	900
Qy	860	TKTPAAATKPEMTTTAKDKITTLERDLRTPPTTTAAAPKMTKETATTTKTESKIIATTTQV	919
Db	901	TKTPAAATKPEMTTTAKDKITTLERDLRTPPTTTAAAPKMTKETATTTKTESKIIATTTQV	960
Qy	920	TSITTTQDTPPKITTLTKITTLAPKVTTTKKTIITTEIMNKPEETAKPKOATNSKATTPK	979
Db	961	TSITTTQDTPPKITTLTKITTLAPKVTTTKKTIITTEIMNKPEETAKPKOATNSKATTPK	1020
Qy	980	PQKPTKAPKXPTSTKKPKTTPVRVKPTTTPTRKMTSTMPELAPTRISAEAMLQTTTRPN	1039
Db	1021	PQKPTKAPKXPTSTKKPKTTPVRVKPTTTPTRKMTSTMPELAPTRISAEAMLQTTTRPN	1080
Qy	1040	QTPNSKULVENVPKSGEDAGGAETTPHMLLRPHVMPPEVTPDMDYLPRVPMQGIINPMLS	1099
Db	1081	QTPNSKULVENVPKSGEDAGGAETTPHMLLRPHVMPPEVTPDMDYLPRVPMQGIINPMLS	1140

```

RESULT 3
US-07-757-022B-2
; Sequence 2, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/07757,022B
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA: US 07/546,114
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-2

Query Match          99.8%; Score 5820.9; DB 4; Len
Best Local Similarity 96.4%; Pred.No.1.3e-169;
Matches 1099; Conservative 0; Mismatches 0; Indel

Qy 1 MAWKTLPTLYLLLLSVFVIQVSSQ-----
Db 1 MAWKTLPTLYLLLLSVFVIQVSSQDLSACRGCGEYSRDTACN
Qy 26 -----ELSCKGRCFESFERGREGCDCAQCKKYDKCPDYESFCA
Db 61 KRVTAEELSCKGRCFESFERGREGCDCAQCKKYDKCPDYESFCA
Qy 80 PPSGASQTIKTTKRSKPPNKKTKKVIIESEETIEHHSVSENGH
Db 121 PPSGASQTIKTTKRSKPPNKKTKKVIIESEETIEHHSVSENGH
Qy 140 KIKSKNSAANRELQKLIKVDKNKKNRTKKPKTPKPPVVDVAGS

```

Db 181 KIKSSKNSAANRELQKLLKLVKDNKNRKKKPTPKPPVVDEAGSLDNGDFKVTTPDTST 240
 Qy 200 TOHNKVSTSPKITTAKPINRPSLSPNSDTSKETSLSLVNKTETVETKETTNNQTSIDG 259
 Db 241 TOHNKVSTSPKITTAKPINRPSLSPNSDTSKETSLSLVNKTETVETKETTNNQTSIDG 300
 Qy 260 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPAS 319
 Db 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPAS 360
 Qy 320 TTPKESPTTIIKSAPTTPKBPAPTTTKSAPTTTKBPAPTTTKBPAPTTTKBP 379
 Db 361 TTPKESPTTIIKSAPTTPKBPAPTTTKSAPTTTKBPAPTTTKBPAPTTTKBP 420
 Qy 380 APTTTKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTK 439
 Db 421 APTTTKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTK 480
 Qy 440 EPAPTAPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPT 499
 Db 481 EPAPTAPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPT 540
 Qy 500 TTKSAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 559
 Db 541 TTKSAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 600
 Qy 560 APTAPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 619
 Db 601 APTAPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPT 660
 Qy 620 PEEAPPTPKAAANPTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPT 679
 Db 661 PEEAPPTPKAAANPTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPT 720
 Qy 680 APTTPKBPAPLAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPT 739
 Db 721 APTTPKBPAPLAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPT 780
 Qy 740 TAPTTKBPAPTTPKKAPKELAPTTTKGPTSTSDKAPTTTKBPAPTTTKBPAPTTPK 799
 Db 781 TAPTTKBPAPTTPKKAPKELAPTTTKGPTSTSDKAPTTTKBPAPTTTKBPAPTTPK 840
 Qy 800 KPAPTTTPPTTSEVSTPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 859
 Db 841 KPAPTTTPPTTSEVSTPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 900
 Qy 860 TKTPAATKPEMTTAKDXTTERDLRTTPTTTPAAPTKEATTTTETKTSKITATTQV 919
 Db 901 TKTPAATKPEMTTAKDXTTERDLRTTPTTTPAAPTKEATTTTETKTSKITATTQV 960
 Qy 920 TSTTTQDTPPKITTLTKTLAPKVTTKITITTEIMNKPETAKPKDRATNSKATTPK 979
 Db 961 TSTTTQDTPPKITTLTKTLAPKVTTKITITTEIMNKPETAKPKDRATNSKATTPK 1020
 Qy 980 POKETKAPKPTSTKPKIMPRVKPKTTPTRKWTSTMPELNPTSRIAEAMQTTTRPN 1039
 Db 1021 POKETKAPKPTSTKPKIMPRVKPKTTPTRKWTSTMPELNPTSRIAEAMQTTTRPN 1080
 Qy 1040 QTPNSKLVEVNPKSGEDAGGAGETPHMLLRPHVPMPEVTPDMDYLPVNPQGIINPMLS 1099
 Db 1081 QTPNSKLVEVNPKSGEDAGGAGETPHMLLRPHVPMPEVTPDMDYLPVNPQGIINPMLS 1140

RESULT 4

US-07-757-022B-62
 ; Sequence 62, Application US/07757022B
 ; Patent No. 6433142
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesner, Thomas G.
 ; APPLICANT: Clark, Stephen C.
 ; APPLICANT: Turner, Katherine
 ; APPLICANT: Hewick, Rodney M.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 CambridgePark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/757,022B
 FILING DATE: 19910910
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Cserr, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851
 INFORMATION FOR SEQ ID NO: 62:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1404 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-757-022B-62

Query Match 99.8%; Score 5820.9; DB 4; Length 1404;
 Best Local Similarity 96.4%; Pred. No. 1.3e-169;
 Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

Qy 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
 Db 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGYSRDATCNDYNCQHYMECCDF 60
 Qy 26 -----ELSCGRCPESFERGECDCDAQCKYDKCCPDYEFCAEVHNPTSPSSSKAP 79
 Db 61 KRVCYTAELSCGRCPESFERGECDCDAQCKYDKCCPDYEFCAEVHNPTSPSSSKAP 120
 Qy 80 PPSGASQIKSTTKSPKPPNKKTKKVIIESEETEEHSVSENESSSSSSSSSSSTI 139
 Db 121 PPSGASQIKSTTKSPKPPNKKTKKVIIESEETEEHSVSENESSSSSSSSSSSTI 180
 Qy 140 KIKSSKNSAANRELQKLLKLVKDNKNRKKKPTPKPPVVDEAGSLDNGDFKVTTPDTST 199
 Db 181 KIKSSKNSAANRELQKLLKLVKDNKNRKKKPTPKPPVVDEAGSLDNGDFKVTTPDTST 240
 Qy 200 TOHNKVSTSPKITTAKPINRPSLSPNSDTSKETSLSLVNKTETVETKETTNNQTSIDG 259
 Db 241 TOHNKVSTSPKITTAKPINRPSLSPNSDTSKETSLSLVNKTETVETKETTNNQTSIDG 300
 Qy 260 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPAS 319
 Db 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPAS 360
 Qy 320 TTPKESPTTIIKSAPTTPKBPAPTTTKSAPTTTKBPAPTTTKBPAPTTTKBPAPT 379

QY 1040 QTPNSKLVNPKSDEAGAGETPHMLLRPHVFMPEVTPDMYLPVFNQGIINPMLS 1099
 Db 1081 QTPNSKLVNPKSDEAGAGETPHMLLRPHVFMPEVTPDMYLPVFNQGIINPMLS 1140

RESULT 6
 US-10-164-595-78
 ; Sequence 78, Application US/10164595
 ; Patent No. 6657054
 ; GENERAL INFORMATION:
 ; APPLICANT: OriGene Technologies, Inc
 ; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
 ; FILE REFERENCE: 1U 103 R1
 ; CURRENT APPLICATION NUMBER: US/10/164,595
 ; CURRENT FILING DATE: 2002-06-10
 ; NUMBER OF SEQ ID NOS: 80
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 78
 ; LENGTH: 1404
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-164-595-78

Query Match 99.4%; Score 5797.9; DB 4; Length 1404;
 Best Local Similarity 96.1%; Pred. No. 6.3e-169;
 Matches 1096; Conservative 0; Mismatches 3; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
 Db 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGCGYSRDATCNCYNOCHNMECCPDF 60

QY 26 -----ELSCRCRCFESPERGECDCDAQCKYDKCCPDYDFCAEVHNPTSPSSKKAP 79
 Db 61 KVCVTAELSCRCRCFESPERGECDCDAQCKYDKCCPDYDFCAEVHNPTSPSSKKAP 120

QY 80 PPSGASQIKTTKESPPNKKTKKVIIEEBITEHSVSENQESSSSSSSSSTI 139
 Db 121 PPSGASQIKTTKRSPPNKKTKKVIIEEBITEHSVSENQESSSSSSSSSTIR 180

QY 140 KIKSSKNAANRELQKLVKNDKNKRTKKPTKPPVWDAGSLDNGDFKVTTPDTST 199
 Db 181 KIKSSKNAANRELQKLVKNDKNKRTKKPTKPPVWDAGSLDNGDFKVTTPDTST 240

QY 200 TQHNKVSPTKITAKPINRPSLPNDSKTSLSLVNKEVETVETKTTNKQTS 259
 Db 241 TQHNKVSPTKITAKPINRPSLPNDSKTSLSLVNKEVETVETKTTNKQTS 300

QY 260 KEKTTSAKETOSIEKTSADKLAPSKVLAKPTPKAETTKGPALTTKPEPTTPKPEPAS 319
 Db 301 KEKTTSAKETOSIEKTSADKLAPSKVLAKPTPKAETTKGPALTTKPEPTTPKPEPAS 360

QY 320 TTPKEPTTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPE 379
 Db 361 TTPKEPTTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPE 420

QY 380 APTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTK 439
 Db 421 APTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTK 480

QY 440 EPAPTAPKPAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPT 499
 Db 481 EPAPTAPKPAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPT 540

QY 500 TTKSAPTTPKPEPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKKP 559
 Db 541 TTKSAPTTPKPEPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKKP 600

QY 560 APTAPKPEAPTTTPKETAPTTPKLTPTTPEKLAFTTPKPAFTTPELAFTTPEEPTPT 619
 Db 601 APTTPKPEAPTTTPKETAPTTPKLTPTTPEKLAFTTPKPAFTTPELAFTTPEEPTPT 660

QY 620 PEEAPPTPKAAANTPKBAPPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTT 679

Db 661 PEEAPPTPKAAANTPKBAPPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTT 720
 QY 680 APTTPKPKAPKELAPTTTKPTSTTSKDPAPTTPKGTAPTTPKBPAPTTPKBPAPTT 739
 Db 721 APTTPKPKAPKELAPTTTKPTSTTSKDPAPTTPKGTAPTTPKBPAPTTPKBPAPTT 780

QY 740 TAPTTLKBPAPTTPKBPAPKELAPTTTKGTSTTSKDPAPTTPKGTAPTTPKBPAPTT 799
 Db 781 TAPTTLKBPAPTTPKBPAPKELAPTTTKGTSTTSKDPAPTTPKGTAPTTPKBPAPTT 840

QY 800 KPAPTTPETPTTSEVSTPTTTKEPTTIHKSPDSTPELSAETPKALENSPKPSPGVT 859
 Db 841 KPAPTTPETPTTSEVSTPTTTKEPTTIHKSPDSTPELSAETPKALENSPKPSPGVT 900

QY 860 TKTPAATKPEMTTTAKDKTTERDLRTTPTTAAAPKMTKETATTTKTESKITATTTQV 919
 Db 901 TKTPAATKPEMTTTAKDKTTERDLRTTPTTAAAPKMTKETATTTKTESKITATTTQV 960

QY 920 TSTTTQDTTPEKLTTLTKTTLAPKVTITTKTITTEIMNKPBEAKPKDRATNSKATTPK 979
 Db 961 TSTTTQDTTPEKLTTLTKTTLAPKVTITTKTITTEIMNKPBEAKPKDRATNSKATTPK 1020

QY 980 PQKPTKAPKPTSTKKPKTMRPRVKPTTPTPRKMTSTMPELNPTSRIAEAMLOTTTRPN 1039
 Db 1021 PQKPTKAPKPTSTKKPKTMRPRVKPTTPTPRKMTSTMPELNPTSRIAEAMLOTTTRPN 1080

QY 1040 QTPNSKLVNPKSDEAGAGETPHMLLRPHVFMPEVTPDMYLPVFNQGIINPMLS 1099
 Db 1081 QTPNSKLVNPKSDEAGAGETPHMLLRPHVFMPEVTPDMYLPVFNQGIINPMLS 1140

RESULT 7
 US-07-757-022B-46
 ; Sequence 46, Application US/07757022B
 ; Patent No. 6433142
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesner, Thomas G.
 ; APPLICANT: Clark, Stephen C.
 ; APPLICANT: Turner, Katherine
 ; APPLICANT: Hewick, Rodney M.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/757,022B
 ; FILING DATE: 19910910
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/546,114
 ; FILING DATE: 29-JUN-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cseert, Luann
 ; REGISTRATION NUMBER: 31,822


```

; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-46

Query Match          96.3%; Score 5617.7; DB 4; Length 1320;
Best Local Similarity 96.1%; Pred. No. 1.8e-163;
Matches 1056; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSCGRCFSPERGECDCDAOCKYDKCCPDYE 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQELSCGRCFSPERGECDCDAOCKYDKCCPDYE 60

QY 61 SFCAEVHNPTSPSSKAPPPSGASQTIKSTTKRSPKPPNKKTKKVIKIESEIIEHSVS 120
DB 61 SFCAEVHNPTSPSSKAPPPSGASQTIKSTTKRSPKPPNKKTKKVIKIESEIIE 115

QY 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKONKQRTKKKTPKPPVDE 180
DB 116 -----VKONKQRTKKKTPKPPVDE 137

QY 181 AGSGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKE 240
DB 138 AGSGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKE 197

QY 241 TVVETKETTITNNKQISTDCKEKTTSKAKETQSIKTSKDLAPTSKVLAKPTPKAETTKG 300
DB 198 TVVETKETTITNNKQISTDCKEKTTSKAKETQSIKTSKDLAPTSKVLAKPTPKAETTKG 257

QY 301 PALTTPKEPTTPKPEASTTKEPTTIKSAPTTPKEPATTTKSAPTTPKPEAPTTP 360
DB 258 PALTTPKEPTTPKPEASTTKEPTTIKSAPTTPKPEAPTTPKSAPTTPKPEAPTTP 317

QY 361 KEPAPTTPKEPATTTKPEATTTKSAPTTPKEPATTPPKKPAPTTPKPEAPTTP 420
DB 318 KEPAPTTPKEPATTTKPEATTTKSAPTTPKEPATTPPKKPAPTTPKPEAPTTP 377

QY 421 TTPKEPATTPKPEATTPKPEATTPKPEATTPKPEATTPKPEATTPKPEATTPKE 480
DB 378 TTPKEPATTPKPEATTPKPEATTPKPEATTPKPEATTPKPEATTPKPEATTPKE 437

QY 481 PAPTTPKSAPTTPKPEATTTKSAPTTPKPEPTTPKPEATTPKPEATTPKPEAPTTP 540
DB 438 PAPTTPKSAPTTPKPEATTTKSAPTTPKPEPTTPKPEATTPKPEATTPKPEAPTTP 497

QY 541 KEPAPTTPKEPATTTKKGAPTAPKPEPATTPKETAAPTTPKXLTPTTPKLAAPTTPKPA 600
DB 498 KEPAPTTPKEPATTTKKGAPTAPKPEPATTPKETAAPTTPKXLTPTTPKLAAPTTPKPA 557

QY 601 PTTPEELAPTTPEEPTTPPEEAPTTPKAAANTPKPEAPTTPKPEAPTTPKPEAPTTP 660
DB 558 PTTPEELAPTTPEEPTTPPEEAPTTPKAAANTPKPEAPTTPKPEAPTTPKPEAPTTP 617

QY 661 KETAPTTPKGAPTTPKPAKPELATTTKGTSTTSKPAAPTTPKGTAPTTP 720
DB 618 KETAPTTPKGAPTTPKPAKPELATTTKGTSTTSKPAAPTTPKGTAPTTP 677

QY 721 PKPEAPTTPKPEAPTTPKGTAPTTPKPAKPELATTTKGTSTTSKPAAPTTPKGTAPTTP 780
DB 678 PKPEAPTTPKPEAPTTPKGTAPTTPKPAKPELATTTKGTSTTSKPAAPTTPKGTAPTTP 737

QY 781 TPKETAPTTPKEPATTPKPEAPTTPBTPPTTSEVSTPTTTKEPTTIHKSPDESIPELS 840
DB 738 TPKETAPTTPKEPATTPKPEAPTTPBTPPTTSEVSTPTTTKEPTTIHKSPDESIPELS 797

QY 841 ABPTPKALENSPKPEGVPTTKTTPAATKPEMTTAKOKTTERDLRTTPEITTAAPKMTKET 900

```

```

RESULT 8
US-07-757-022B-60
; Sequence 60, Application US/07/57022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Csezz, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-60

```

Query Match 96.3%; Score 5617.7; DB 4; Length 1320;
 Best Local Similarity 96.1%; Pred. No. 1.8e-163;
 Matches 1056; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQBLSCGRCFSEFGRGECDDAQCCKYDKCCPDYE 60
 DB 1 MAWKTLPIYLLLLSVFVIQVSSQBLSCGRCFSEFGRGECDDAQCCKYDKCCPDYE 60

QY 61 SFCAEVHNTPSPSSKAPPPSGASQTIKSTTKRSPKPNKKTKKVIIESEBITEHSVS 120
 DB 61 SFCAEVHNTPSPSSKAPPPSGASQTIKSTTKRSPKPNKKTKKVIIESEBITE- ---- 115

QY 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKKLVKONKKKRTKKKPTPKPPVVDE 180
 DB 116 -----VKONKKKRTKKKPTPKPPVVDE 137

QY 181 AGSGLDNGDFKVTTPDTSTTOHNVSTSPKITTAKPINPRPSLPNSDTSKETSILTVNKE 240
 DB 138 AGSGLDNGDFKVTTPDTSTTOHNVSTSPKITTAKPINPRPSLPNSDTSKETSILTVNKE 197

QY 241 TTVEIKETITTKQSTDKGKETTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKG 300
 DB 198 TTVEIKETITTKQSTDKGKETTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKG 257

QY 301 PALTTTKEPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSAPTTPKEPAPTTT 360
 DB 258 PALTTTKEPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSAPTTPKEPAPTTT 317

QY 361 KEPAATTPKEPAPTTTKEPAPTTTTSAPTTPKEPAPTTTTSAPTTPKEPAPTTTKEPTT 420
 DB 318 KEPAATTPKEPAPTTTKEPAPTTTTSAPTTPKEPAPTTTTSAPTTPKEPAPTTTKEPTT 377

QY 421 TTPKEPAPTTTKEPAPTTTKEPAPTTAPKAPAPTTTKEPAPTTTKEPAPTTTKEPSPTTKE 480
 DB 378 TTPKEPAPTTTKEPAPTTTKEPAPTTAPKAPAPTTTKEPAPTTTKEPAPTTTKEPSPTTKE 437

QY 481 PAPTITKSGAPTTTKEPAPTTTTSAPTTPKESPPTTKEPAPTTTKEPAPTTTKEPAPTTT 540
 DB 438 PAPTITKSGAPTTTKEPAPTTTTSAPTTPKESPPTTKEPAPTTTKEPAPTTTKEPAPTTT 497

QY 541 KEPAATTPKEPAPTTTKEPAPTTAPKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 600
 DB 498 KEPAATTPKEPAPTTTKEPAPTTAPKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 557

QY 601 PTTPEELAPTTPEEPTTTPKEPAPTTTTSAPTTPKAAPTTPKEPAPTTTKEPAPTTTKEPAPTTT 660
 DB 558 PTTPEELAPTTPEEPTTTPKEPAPTTTTSAPTTPKAAPTTPKEPAPTTTKEPAPTTTKEPAPTTT 617

QY 661 KETAPTTKGTAPTTTKEPAPTTTTPKAPAPKELAPTTTKEPTTSTSDKAPAPTTTKEGTA 720
 DB 618 KETAPTTKGTAPTTTKEPAPTTTTPKAPAPKELAPTTTKEPTTSTSDKAPAPTTTKEGTA 677

QY 721 PKEPAPTTTKEPAPTTTKEGTAAPTTTKEPAPTTTTPKAPAPKELAPTTTKEGTAAPTTTSDK 780
 DB 678 PKEPAPTTTKEPAPTTTKEGTAAPTTTKEPAPTTTTPKAPAPKELAPTTTKEGTAAPTTTSDK 737

QY 781 TPKETAPTTTKEPAPTTTKEPAPTTTTPETPTTSEVSTPTTKEPAPTTTTPKAPAPKELAP 840
 DB 738 TPKETAPTTTKEPAPTTTKEPAPTTTTPETPTTSEVSTPTTKEPAPTTTTPKAPAPKELAP 797

QY 841 AEPTPKALENSPKPGVPTTKTAPATPEMTTAKDKTTERDLTTPETTTAAAPKMTKET 900
 DB 798 AEPTPKALENSPKPGVPTTKTAPATPEMTTAKDKTTERDLTTPETTTAAAPKMTKET 857

QY 901 ATTTEKTTESKITATTTQVSTTTQDPTTTPFKITLTKTTTLAPKVTITTKKTIITTEIMNKP 960
 DB 858 ATTTEKTTESKITATTTQVSTTTQDPTTTPFKITLTKTTTLAPKVTITTKKTIITTEIMNKP 917

QY 961 EETAKPKDRATNSKATTPKPKOPTKAPKPTSTKPKTMTBRVRKPTTTPRKMSTMP 1020
 DB 918 EETAKPKDRATNSKATTPKPKOPTKAPKPTSTKPKTMTBRVRKPTTTPRKMSTMP 977

QY 1021 INPSTRIAEAMLOTTTRPNQTPNSKLVVNPVKSEDAGGAEGETPHMLLRPHVFMPEVTPD 1080
 DB 978 INPSTRIAEAMLOTTTRPNQTPNSKLVVNPVKSEDAGGAEGETPHMLLRPHVFMPEVTPD 1037

QY 1081 MDYLPVRVNOGIILINPMLS 1099
 DB 1038 MDYLPVRVNOGIILINPMLS 1056

RESULT 9
 US-10-164-595-58
 ; Sequence 58; Application US/10164595
 ; Patent No. 6657054
 ; GENERAL INFORMATION:
 ; APPLICANT: Origene Technologies, Inc
 ; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
 ; FILE REFERENCE: IU 103 R1
 ; CURRENT APPLICATION NUMBER: US/10/164,595
 ; CURRENT FILING DATE: 2002-06-10
 ; NUMBER OF SEQ ID NOS: 80
 ; SOFTWARE: Patencin version 3.1
 ; SEQ ID NO 58
 ; LENGTH: 1320
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-164-595-58

Query Match 96.1%; Score 5608.7; DB 4; Length 1320;
 Best Local Similarity 95.9%; Pred. No. 3.4e-163;
 Matches 1054; Conservative 0; Mismatches 2; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQBLSCGRCFSEFGRGECDDAQCCKYDKCCPDYE 60
 DB 1 MAWKTLPIYLLLLSVFVIQVSSQBLSCGRCFSEFGRGECDDAQCCKYDKCCPDYE 60

QY 61 SFCAEVHNTPSPSSKAPPPSGASQTIKSTTKRSPKPNKKTKKVIIESEBITEHSVS 120
 DB 61 SFCAEVHNTPSPSSKAPPPSGASQTIKSTTKRSPKPNKKTKKVIIESEBITE- ---- 115

QY 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKKLVKONKKKRTKKKPTPKPPVVDE 180
 DB 116 -----VKONKKKRTKKKPTPKPPVVDE 137

QY 181 AGSGLDNGDFKVTTPDTSTTOHNVSTSPKITTAKPINPRPSLPNSDTSKETSILTVNKE 240
 DB 138 AGSGLDNGDFKVTTPDTSTTOHNVSTSPKITTAKPINPRPSLPNSDTSKETSILTVNKE 197

QY 241 TTVEIKETITTKQSTDKGKETTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKG 300
 DB 198 TTVEIKETITTKQSTDKGKETTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKG 257

QY 301 PALTTTKEPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSAPTTPKEPAPTTT 360
 DB 258 PALTTTKEPTTTPKEPASTTTPKEPTTTIKSAPTTPKEPAPTTTTSAPTTPKEPAPTTT 317

QY 361 KEPAATTPKEPAPTTTKEPAPTTTTSAPTTPKEPAPTTTTSAPTTPKEPAPTTTKEPTT 420
 DB 318 KEPAATTPKEPAPTTTKEPAPTTTTSAPTTPKEPAPTTTTSAPTTPKEPAPTTTKEPTT 377

QY 421 TTPKEPAPTTTKEPAPTTTKEPAPTTAPKAPAPTTTKEPAPTTTKEPAPTTTKEPSPTTKE 480
 DB 378 TTPKEPAPTTTKEPAPTTTKEPAPTTAPKAPAPTTTKEPAPTTTKEPAPTTTKEPSPTTKE 437

QY 481 PAPTITKSGAPTTTKEPAPTTTTSAPTTPKESPPTTKEPAPTTTKEPAPTTTKEPAPTTT 540
 DB 438 PAPTITKSGAPTTTKEPAPTTTTSAPTTPKESPPTTKEPAPTTTKEPAPTTTKEPAPTTT 497

QY 541 KEPAATTPKEPAPTTTKEPAPTTAPKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 600
 DB 498 KEPAATTPKEPAPTTTKEPAPTTAPKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 557

QY 601 PTTPEELAPTTPEEPTTTPKEPAPTTTTSAPTTPKAAPTTPKEPAPTTTKEPAPTTTKEPAPTTT 660
 DB 558 PTTPEELAPTTPEEPTTTPKEPAPTTTTSAPTTPKAAPTTPKEPAPTTTKEPAPTTTKEPAPTTT 617

QY 661 KETAPTTKGTAPTTTKEPAPTTTTPKAPAPKELAPTTTKEPTTSTSDKAPAPTTTKEGTA 720
 DB 618 KETAPTTKGTAPTTTKEPAPTTTTPKAPAPKELAPTTTKEPTTSTSDKAPAPTTTKEGTA 677

QY 721 PKEPAPTTTKEPAPTTTKEGTAAPTTTKEPAPTTTTPKAPAPKELAPTTTKEGTAAPTTTSDK 780
 DB 678 PKEPAPTTTKEPAPTTTKEGTAAPTTTKEPAPTTTTPKAPAPKELAPTTTKEGTAAPTTTSDK 737

QY 781 TPKETAPTTTKEPAPTTTKEPAPTTTTPETPTTSEVSTPTTKEPAPTTTTPKAPAPKELAP 840
 DB 738 TPKETAPTTTKEPAPTTTKEPAPTTTTPETPTTSEVSTPTTKEPAPTTTTPKAPAPKELAP 797

QY 841 AEPTPKALENSPKPGVPTTKTAPATPEMTTAKDKTTERDLTTPETTTAAAPKMTKET 900
 DB 798 AEPTPKALENSPKPGVPTTKTAPATPEMTTAKDKTTERDLTTPETTTAAAPKMTKET 857

QY 901 ATTTEKTTESKITATTTQVSTTTQDPTTTPFKITLTKTTTLAPKVTITTKKTIITTEIMNKP 960
 DB 858 ATTTEKTTESKITATTTQVSTTTQDPTTTPFKITLTKTTTLAPKVTITTKKTIITTEIMNKP 917

QY 961 EETAKPKDRATNSKATTPKPKOPTKAPKPTSTKPKTMTBRVRKPTTTPRKMSTMP 1020
 DB 918 EETAKPKDRATNSKATTPKPKOPTKAPKPTSTKPKTMTBRVRKPTTTPRKMSTMP 977

558 PTTPEELAPTTPEETPTTPEBPAPTTPKAAAPNTKEBPAPTTKEBPAPTTKEBPAPTT 617
661 KETAPTTTKEGAPTTTKEBPAPTTKPKAPKELAPTTTKEPTSTSDKEPAPTTKGTAPTT 720
618 KETAPTTTKEGAPTTTKEBPAPTTKPKAPKELAPTTTKEPTSTTCDKBPAPTTKGTAPTT 677
721 KPEAPTTTKEBPAPTTTKEGAPTTTKEBPAPTTTKEBPAPTTTKEPTSTTSDKBPAPTT 780
678 KPEAPTTTKEBPAPTTTKEGAPTTTKEBPAPTTTKEBPAPTTTKEPTSTTSDKBPAPTT 737
781 TPEKAPTTTKEBPAPTTTKEGAPTTTKEBPAPTTTKEBPAPTTTKEPTSTTSDKBPAPTT 840
738 TPEKAPTTTKEBPAPTTTKEGAPTTTKEBPAPTTTKEBPAPTTTKEPTSTTSDKBPAPTT 797
841 APTTPEKALENSKEGAPTTTKEGAPTTTKEBPAPTTTKEBPAPTTTKEPTSTTSDKBPAPTT 900
798 APTTPEKALENSKEGAPTTTKEGAPTTTKEBPAPTTTKEBPAPTTTKEPTSTTSDKBPAPTT 857
901 ATTTEKTESKITATTQTQVSTTTQDTPPKTTLKTTTLAPKVTTKKLTITTEIMNKP 960
858 ATTTEKTESKITATTQTQVSTTTQDTPPKTTLKTTTLAPKVTTKKLTITTEIMNKP 917
961 ESTAKPKDRAATSKATTPKPKQPTKAPKPKPTSTKPKTMVRKPKTTTTPRKMSTWPE 1020
918 ESTAKPKDRAATSKATTPKPKQPTKAPKPKPTSTKPKTMVRKPKTTTTPRKMSTWPE 977
1021 LNPTSRIBAEMLQTTTRPNQTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 1080
978 LNPTSRIBAEMLQTTTRPNQTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 1037
1081 MDYLPVPRVNOGIIINPMLS 1099
1038 MDYLPVPRVNOGIIINPMLS 1056

RESULT 10
US-07-757-0228-40
Sequence 40; Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1361 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-0228-40

Query Match 96.0%; Score 5603.6; DB 4; Length 1361;
Best Local Similarity 92.6%; Pred. No. 5e-163;
Matches 1056; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

Qy 1 MANKTLPIYLLLSVFIQVSSQ----- 25
Db 1 MANKTLPIYLLLSVFIQVSSQ----- 60
Qy 26 -----ELSCCKGRCFESFERGECDDAQCCKYDKCCPDYESFCAEVHNFTSPSSKKAP 79
Db 61 KRVCATLSCCKGRCFESFERGECDDAQCCKYDKCCPDYESFCAEVHNFTSPSSKKAP 120
Qy 80 PPGASQTIKSTTKRSPKPNKKTKKVISEBITEHVSSESSSSSSSSSSSSSTIW 139
Db 121 PPGASQTIKSTTKRSPKPNKKTKKVISEBITE----- 156
Qy 140 KIKSKNSAANRELOKLLKVKDNKKNRTKKKPTKPPVVDAGSLDNGDFKVTPTST 199
Db 157 -----VKDNKKNRTKKKPTKPPVVDAGSLDNGDFKVTPTST 197
Qy 200 TOHNKVSTSPKITTAKPINPRPSLPNSDTSKETSITVNETTETTKTNNKQSTGD 259
Db 198 TOHNKVSTSPKITTAKPINPRPSLPNSDTSKETSITVNETTETTKTNNKQSTGD 257
Qy 260 KEKTSKAKETQSIKTSKADLAPTSKVLAKPTKPAETTTKGPALTTKEPTTTTKEPAS 319
Db 258 KEKTSKAKETQSIKTSKADLAPTSKVLAKPTKPAETTTKGPALTTKEPTTTTKEPAS 317
Qy 320 TTPKEPTTTTISAPTTTKEBPAPTTTISAPTTTKEBPAPTTTKEBPAPTTTKEP 379
Db 318 TTPKEPTTTTISAPTTTKEBPAPTTTISAPTTTKEBPAPTTTKEBPAPTTTKEP 377
Qy 380 APTTTSKAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTT 439
Db 378 APTTTSKAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTT 437
Qy 440 EPAPTAPKPKAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTT 499
Db 438 EPAPTAPKPKAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTT 497
Qy 500 TTKSAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTT 559
Db 498 TTKSAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTT 557
Qy 560 APTAPKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTT 619
Db 558 APTAPKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTT 617
Qy 620 PEEAPTTTPKAAAPNTPKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTT 679
Db 618 PEEAPTTTPKAAAPNTPKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTT 677
Qy 680 APTTPKKAPKELAPTTTKEPTSTSDKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTT 739
Db 678 APTTPKKAPKELAPTTTKEPTSTSDKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTT 737
Qy 740 TAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTT 799

Db 738 TAPTTLKAPAPTPPKKPAKELAPTTTKGPTSTSDKPAPTTKEATPTPKAPAPTPK 797
Qy 800 KPAPPTPEPTPTTSVSTPTTKETTTTHKSPDESTPELSAEPKALENSPKPGVPT 859
Db 798 KPAPPTPEPTPTTSVSTPTTKETTTTHKSPDESTPELSAEPKALENSPKPGVPT 857
Qy 860 TKTPAATKPEMTTAKDKTTERDLRTTPTTTAAAPKWKETATTTTEKTESKITATTQV 919
Db 858 TKTPAATKPEMTTAKDKTTERDLRTTPTTTAAAPKWKETATTTTEKTESKITATTQV 917
Qy 920 TSTTTQDTTFFKLTTLKTTLLAPKVTTKKTTTTEIMNKPBTAKPDATNSKATTPK 979
Db 918 TSTTTQDTTFFKLTTLKTTLLAPKVTTKKTTTTEIMNKPBTAKPDATNSKATTPK 977
Qy 980 PQKPTKAPKPTSTTKPKTMVPRKPTTPTPKMTSTWPELNPSTSRIAEAMLTQTTREN 1039
Db 978 PQKPTKAPKPTSTTKPKTMVPRKPTTPTPKMTSTWPELNPSTSRIAEAMLTQTTREN 1037
Qy 1040 QTPNSKLVEVNPKSEDAGAGETHMLLRPHVFMPEVTPDMDYLPRVFNQIINPMLS 1099
Db 1038 QTPNSKLVEVNPKSEDAGAGETHMLLRPHVFMPEVTPDMDYLPRVFNQIINPMLS 1097

RESULT 11
US-07-757-022B-58
; Sequence 58, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turney, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1049 amino acids

; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-58
Query Match 95.3%; Score 5561; DB 4; Length 1049;
Best Local Similarity 95.5%; Pred. No. 7.1e-162;
Matches 1049; Conservative 0; Mismatches 0; Indels 50; Gaps 1;
Qy 1 MAWKTLPIYLLLLSVFVIOQVSSQBELSKGRGCFESFERGECDDAQQKDYKCCPDYE 60
Db 1 MAWKTLPIYLLLLSVFVIOQVSSQBELSKGRGCFESFERGECDDAQQKDYKCCPDYE 60
Qy 61 SFCAEVENPTSPSSKKAPPPSGASQTIKSTTKRSPPKPKKTKKVISEETESHSVS 120
Db 61 SFCA-----EHSVS 70
Qy 121 ENOESSSSSSSSSSSTIWKIKSSKNSAANRELQKLVKDNKQNTKKKPTPKPVWDE 180
Db 71 ENOESSSSSSSSSSSTIWKIKSSKNSAANRELQKLVKDNKQNTKKKPTPKPVWDE 130
Qy 181 AGSGLDNGDFKVTTPDTSTTOHNVKSTSPKITTAKPINRPSLPNSDTSKETSLTWKE 240
Db 131 AGSGLDNGDFKVTTPDTSTTOHNVKSTSPKITTAKPINRPSLPNSDTSKETSLTWKE 190
Qy 241 TTVETKETTNNKQSTDGKEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTKG 300
Db 191 TTVETKETTNNKQSTDGKEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTKG 250
Qy 301 PALTPKEPTPTTPKEPASTTKEPTTTIKSAPTTKEPAPTTTKSAPTTKEPAPTTT 360
Db 251 PALTPKEPTPTTPKEPASTTKEPTTTIKSAPTTKEPAPTTTKSAPTTKEPAPTTT 310
Qy 361 KEPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 420
Db 311 KEPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 370
Qy 421 TTPKEPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 480
Db 371 TTPKEPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 430
Qy 481 PAPTITKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 540
Db 431 PAPTITKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 490
Qy 541 KEPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 600
Db 491 KEPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 550
Qy 601 PTTPELAPTTPEEPTPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 660
Db 551 PTTPELAPTTPEEPTPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 610
Qy 661 KETAPTTKGTAPTTLKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 720
Db 611 KETAPTTKGTAPTTLKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 670
Qy 721 PKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 780
Db 671 PKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 730
Qy 781 TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 840
Db 731 TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 790
Qy 841 AEPKALENSPKPGVPTTKTTPAAKPEMTTAKDKTTERDLRTTPTTTAAAPKWKET 900
Db 791 AEPKALENSPKPGVPTTKTTPAAKPEMTTAKDKTTERDLRTTPTTTAAAPKWKET 850
Qy 901 ATTTTEKTESKITATTQVSTTTQDTTTPFKITTLKTTLLAPKVTTKKTTTTEIMNKP 960
Db 851 ATTTTEKTESKITATTQVSTTTQDTTTPFKITTLKTTLLAPKVTTKKTTTTEIMNKP 910

RESULT 13
 US-07-757-022B-48
 ; Sequence 48, Application US/07757022B
 ; Patent No. 6433142
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesner, Thomas G.
 ; APPLICANT: Clark, Stephen C.
 ; APPLICANT: Turner, Katherine
 ; APPLICANT: Hewick, Rodney M.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/757,022B
 ; FILING DATE: 19910910
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/546,114
 ; FILING DATE: 29-JUN-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cseri, Luann
 ; REGISTRATION NUMBER: 31,822
 ; REFERENCE/DOCKET NUMBER: GI 5190
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)876-1170
 ; TELEFAX: (617)876-5851
 ; INFORMATION FOR SEQ ID NO: 48:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1354 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-07-757-022B-48

Query Match 95.1%; Score 5546.9; DB 4; Length 1354;
 Best Local Similarity 92.0%; Pred. No. 2.6e-161;
 Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;
 QY 1 MAWTLPIYLLLLSVFVIQVSSQ----- 25
 DB 1 MAWTLPIYLLLLSVFVIQVSSQ----- 25
 QY 26 -----ELSCKGRCFESFERGECDDAOCKYDKCCPDYEFCAEVHNPTSPSSKKAP 79
 DB 61 KRVCTAELSCKGRCFESFERGECDDAOCKYDKCCPDYEFCA----- 105
 QY 80 PPSGASQTIKSTTKRSPKPNKXKTKVIBSEITEHSVSENOESSSSSSSSSTI 139
 DB 106 -----EEHSVSENOESSSSSSSSSSSTI 130
 QY 140 KIKSSKNSAANRELQKLVKNDKNKNTKKKPTPKPPVWDEAGSLDNGDFKVTTPDTST 199

Db 131 KIKSSKNSAANRELQKLVKNDKNKNTKKKPTPKPPVWDEAGSLDNGDFKVTTPDTST 190
 QY 200 TQHNKVSTSPKITTAKPINRPSLPPNSDTSKETSJLVNKTETVETKETTNTKOTSTDG 259
 Db 191 TQHNKVSTSPKITTAKPINRPSLPPNSDTSKETSJLVNKTETVETKETTNTKOTSTDG 250
 QY 260 KEKTTSAKTSQTSKTSKADLAPTSKVLAKPPTPKAETTTKGPAETTPKEPTTTKEPAS 319
 Db 251 KEKTTSAKTSQTSKTSKADLAPTSKVLAKPPTPKAETTTKGPAETTPKEPTTTKEPAS 310
 QY 320 TTPKEPTTTIKSAPTTPEKAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTK 379
 Db 311 TTPKEPTTTIKSAPTTPEKAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTK 370
 QY 380 APTTKSAPTTPEKAPTTPKKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTT 439
 Db 371 APTTKSAPTTPEKAPTTPKKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTT 430
 QY 440 EPAPTAPKKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTT 499
 Db 431 EPAPTAPKKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTT 490
 QY 500 TTKSAPTTPEKSPPTTKKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEK 559
 Db 491 TTKSAPTTPEKSPPTTKKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEK 550
 QY 560 APTAPKEPAPTTPKETAPTTPKKLTPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEK 619
 Db 551 APTAPKEPAPTTPKETAPTTPKKLTPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEK 610
 QY 620 PEEPAPTTPKAAAPNTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEK 679
 Db 611 PEEPAPTTPKAAAPNTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEK 670
 QY 680 APTTPKKAPKELAPTTTKETPTSTSDKAPTTPKGTAPTTPKGTAPTTPKGTAPTTPKGT 739
 Db 671 APTTPKKAPKELAPTTTKETPTSTSDKAPTTPKGTAPTTPKGTAPTTPKGTAPTTPKGT 730
 QY 740 TAPTTKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPTTPKETAPTTPKETAPTTPKETAPTT 799
 Db 731 TAPTTKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPTTPKETAPTTPKETAPTTPKETAPTT 790
 QY 800 KPAPTTPEPTTSEVSTPTTKETPTTHKSPDESTPELSAETPKALENSPKEPGVPT 859
 Db 791 KPAPTTPEPTTSEVSTPTTKETPTTHKSPDESTPELSAETPKALENSPKEPGVPT 850
 QY 860 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETATTTESKITTATTTQV 919
 Db 851 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETATTTESKITTATTTQV 910
 QY 920 TSTTTQDTPPKITTLKTTTLAPKVTTKKITTTEIMNKBEETAKPKDRATNSKATTPK 979
 Db 911 TSTTTQDTPPKITTLKTTTLAPKVTTKKITTTEIMNKBEETAKPKDRATNSKATTPK 970
 QY 980 POKPTKAPKPTSTKKPKTMRVRKPTTPTPKMTSTMPKNTSRLAEAMLOTTTRPN 1039
 Db 971 POKPTKAPKPTSTKKPKTMRVRKPTTPTPKMTSTMPKNTSRLAEAMLOTTTRPN 1030
 QY 1040 QTPNSKLVENPKSBDAGAGETPHMLLRPHVFMPEVTDMYDLPRVFNQGIINPMLS 1099
 Db 1031 QTPNSKLVENPKSBDAGAGETPHMLLRPHVFMPEVTDMYDLPRVFNQGIINPMLS 1090

RESULT 14
 US-07-757-022B-50
 ; Sequence 50, Application US/07757022B
 ; Patent No. 6433142
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesner, Thomas G.
 ; APPLICANT: Clark, Stephen C.
 ; APPLICANT: Turner, Katherine
 ; APPLICANT: Hewick, Rodney M.

;; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
;; NUMBER OF SEQUENCES: 143
;; CORRESPONDENCE ADDRESS:
;; ADDRESSES: Genetics Institute, Inc.
;; STREET: 87 CambridgePark Drive
;; CITY: Cambridge
;; STATE: Massachusetts
;; COUNTRY: U.S.A.
;; ZIP: 02140
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: IBM PC compatible
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US 07/757,022B
;; FILING DATE: 19910910
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/643,502
;; FILING DATE: 18-JAN-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/546,114
;; FILING DATE: 29-JUN-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/457,196
;; FILING DATE: 29-DEC-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/390,901
;; FILING DATE: 08-AUG-1989
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Cserr, Luann
;; REGISTRATION NUMBER: 31,922
;; REFERENCE/DOCKET NUMBER: GI 5190
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617) 876-1170
;; TELEFAX: (617) 876-5851
;; INFORMATION FOR SEQ ID NO: 50:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1314 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-07-757-022B-50

Query Match 93.2%; Score 5437.9; DB 4; Length 1314;
Best Local Similarity 93.5%; Pred. No. 5.3e-158;
Matches 1028; Conservative 7; Mismatches 14; Indels 51; Gaps 2;

QY 1 MAWKTLPIYLLLSVFNVCQVSSQEL-SCKGRCEFSFERGECDCDAQCKYDKCCPDY 59
DB 1 MAWKTLPIYLLLSVFNVCQVSSQELSSCAGRGEGYSDATCNCQYHNECCPDF 60
QY 60 EGFCAEVHNPSTPPSSKKAPPPSGASQTIKSTTKRSPKPPNKKTKKVIESEBITEHSV 119
DB 61 KRVC-----TAHSV 70
QY 120 SENQSSSSSSSSSSSTTWIKSSKNSAANRELQKLVKDNKNRTKKKTPKPPVD 179
DB 71 SENQSSSSSSSSSSSTTWIKSSKNSAANRELQKLVKDNKNRTKKKTPKPPVD 130
QY 180 EAGSGLDNGDFKVTTPDTSTQHNKYSTGPKITAKPINRPSLPNSDTSKETSILVKN 239
DB 131 EAGSGLDNGDFKVTTPDTSTQHNKYSTGPKITAKPINRPSLPNSDTSKETSILVKN 190
QY 240 ETVETKETTNNKQTSQKETTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTK 299
DB 191 ETVETKETTNNKQTSQKETTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTK 250
QY 300 GPALTTPKPTTPPKPEASTTPKEPTPTTKSAPTTPKPEAPTTTKSAPTTPKPEAPT 359
DB 251 GPALTTPKPTTPPKPEASTTPKEPTPTTKSAPTTPKPEAPTTTKSAPTTPKPEAPT 310

QY 360 TKPEAPTTKEPAPTTTKPEAPTTTKSAPTTTKPEAPTTTKKPEAPTTTKPEAPTTKEPT 419
DB 311 TKPEAPTTKEPAPTTTKPEAPTTTKSAPTTTKPEAPTTTKKPEAPTTTKPEAPTTKEPT 370
QY 420 PTTKPEAPTTKEPAPTTTKPEAPTTAPKAPPTTKPEAPTTTKPEAPTTTKPEAPTTKEPT 479
DB 371 PTTKPEAPTTKEPAPTTTKPEAPTTAPKAPPTTKPEAPTTTKPEAPTTTKPEAPTTKEPT 430
QY 480 EPAPTTTKSAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 539
DB 431 EPAPTTTKSAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 490
QY 540 PKPEAPTTKEPAPTTTKKPEAPTTAPKAPPTTKPEAPTTTKPEAPTTTKPEAPTTKEPT 599
DB 491 PKPEAPTTKEPAPTTTKKPEAPTTAPKAPPTTKPEAPTTTKPEAPTTTKPEAPTTKEPT 550
QY 600 APTTPEELAPTTPEEPPTTPPEAPTTPKAAAPNTPKPEAPTTTKPEAPTTTKPEAPTT 659
DB 551 APTTPEELAPTTPEEPPTTPPEAPTTPKAAAPNTPKPEAPTTTKPEAPTTTKPEAPTT 610
QY 660 PKETAPTTPKGTAPTTTKPEAPTTPKKAPKELAPTTTKPEPTSTTSKPEAPTTPKGTAPT 719
DB 611 PKETAPTTPKGTAPTTTKPEAPTTPKKAPKELAPTTTKPEPTSTTSKPEAPTTPKGTAPT 670
QY 720 TKPEAPTTKEPAPTTTKPEAPTTTKPEAPTTTKKPEAPTTTKKPEAPTTTKGPTSTTSKPEAP 779
DB 671 TKPEAPTTKEPAPTTTKPEAPTTTKPEAPTTTKKPEAPTTTKKPEAPTTTKGPTSTTSKPEAP 730
QY 780 TTPKETAPTTPKPEAPTTPKKPEAPTTPEPTPPPTTSEVSTPTTKEPTTIHKSPESTPEL 839
DB 731 TTPKETAPTTPKPEAPTTPKKPEAPTTPEPTPPPTTSEVSTPTTKEPTTIHKSPESTPEL 790
QY 840 SAEPTPKALENSPKPEGVPTTKPAATKPEMTTAKDKITERDLRTTPEPTTAAPKMTKE 899
DB 791 SAEPTPKALENSPKPEGVPTTKPAATKPEMTTAKDKITERDLRTTPEPTTAAPKMTKE 850
QY 900 TATTEKTESKITATTQVSTTTQDTPFKITLTKTTLAPKVITTKTITTEIMNK 959
DB 851 TATTEKTESKITATTQVSTTTQDTPFKITLTKTTLAPKVITTKTITTEIMNK 910
QY 960 PEETAKPKDRATNSKATTPKQPKTKAPKPTSTTKPKTWPVRVKPTTTPRKMSTMP 1019
DB 911 PEETAKPKDRATNSKATTPKQPKTKAPKPTSTTKPKTWPVRVKPTTTPRKMSTMP 970
QY 1020 ELNPTSRIAEAMLOTTTRPNQNTNSKLVEYNPKSEDAGGAEGETPHMLLRPHVFMPEVTP 1079
DB 971 ELNPTSRIAEAMLOTTTRPNQNTNSKLVEYNPKSEDAGGAEGETPHMLLRPHVFMPEVTP 1030
QY 1080 DMDYLPRVNPQGIINPMLS 1099
DB 1031 DMDYLPRVNPQGIINPMLS 1050
RESULT 15
US-07-757-022B-74
; Sequence 74, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICANT: PatentIn Release #1.0, Version #1.25
;   APPLICATION NUMBER: US/07/757,022B
;   FILING DATE: 19910910
;   CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/643,502
;     FILING DATE: 18-JAN-1991
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/546,114
;     FILING DATE: 29-JUN-1990
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/457,196
;     FILING DATE: 29-DEC-1989
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/390,901
;     FILING DATE: 08-AUG-1989
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Cseri, Luann
;     REGISTRATION NUMBER: 31,822
;     REFERENCE/DOCKET NUMBER: GI 5190
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (617)876-1170
;       TELEFAX: (617)876-5851
;     INFORMATION FOR SEQ ID NO: 74:
;       SEQUENCE CHARACTERISTICS:
;         LENGTH: 1038 amino acids
;         TYPE: AMINO ACID
;         TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   US-07-757-022B-74

```

```

Query Match          91.8%; Score 5353.7; DB 4; Length 1038;
Best Local Similarity 91.5%; Pred. No. 1.4e-155;
Matches 1006; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSGELSCGRCFESFERGECDCDAQCKKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLLSVFVIQVSSGELSCGRCFESFERGECDCDAQCKKYDKCCPDYE 60
QY 61 SFCARVHNPTSPSSKAPPSPGASQTIKSTTKSPKPPNKKTKYVIESEBITEHSVS 120
Db 61 SFCAR----- 65
QY 121 ENQESSSSSSSSSSSTIWKIKSKNSAANRELOKLVKONKNRTKKKPTPKPPVYDE 180
Db 66 -----VKONKNRTKKKPTPKPPVYDE 87
QY 181 AGSLGNDGFKVTTDTSTTOHNVSTSPKITTAKPINRPSLPPNSDTSKESLTVNKE 240
Db 88 AGSLGNDGFKVTTDTSTTOHNVSTSPKITTAKPINRPSLPPNSDTSKESLTVNKE 147
QY 241 TTVEKETTNTKOTSDGKEKTTSAKETQSIKTSKADLAPTSKVLAKPTKPAEITTKG 300
Db 148 TTVEKETTNTKOTSDGKEKTTSAKETQSIKTSKADLAPTSKVLAKPTKPAEITTKG 207
QY 301 PALTTPKEPTTTKEPASITTPKEPTTTIKSAPTTKPEAPTTKSAPTTPKEAPTTT 360
Db 208 PALTTPKEPTTTKEPASITTPKEPTTTIKSAPTTKPEAPTTKSAPTTPKEAPTTT 267
QY 361 KEPAITTPKEAPTTTKEPASITTPKEAPTTKPEAPTTKPEAPTTKPEAPTTKEPT 420
Db 268 KEPAITTPKEAPTTTKEPASITTPKEAPTTKPEAPTTKPEAPTTKPEAPTTKEPT 327
QY 421 TTPKEAPTTKEAPTTTKEPASITTPKEAPTTKPEAPTTKPEAPTTTKEPSPTTKE 480
Db 328 TTPKEAPTTKEAPTTTKEPASITTPKEAPTTKPEAPTTTKEPSPTTKE 387
QY 481 PAPTTTKEAPTTTKEPASITTPKEAPTTTKEPASITTPKEAPTTTKEPASITTPKE 540
Db 388 PAPTTTKEAPTTTKEPASITTPKEAPTTTKEPASITTPKEAPTTTKEPASITTPKE 447

```

```

QY 541 KEPAITTPKEAPTTTKEPASITTPKEAPTTKPEAPTTKPEAPTTTKEPASITTPKEPA 600
Db 448 KEPAITTPKEAPTTTKEPASITTPKEAPTTKPEAPTTKPEAPTTTKEPASITTPKEPA 507
QY 601 PTTPEELAPTTPEEPTTTPEEPTTTPEEPTTTPEEPTTTPEEPTTTPEEPTTTPEEPTTT 660
Db 508 PTTPEELAPTTPEEPTTTPEEPTTTPEEPTTTPEEPTTTPEEPTTTPEEPTTTPEEPTTT 567
QY 661 KETAPTTPKGTAPTTTLKSPAPTTPKKPAKELAPTTTKEPTTTTSDKPAPTTKGTAPTT 720
Db 568 KETAPTTPKGTAPTTTLKSPAPTTPKKPAKELAPTTTKEPTTTTSDKPAPTTKGTAPTT 627
QY 721 PKEPAPTTKEAPTTPKGTAPTTTLKSPAPTTPKKPAKELAPTTTKEPTTTTSDKPAPTTKGTAPTT 780
Db 628 PKEPAPTTKEAPTTPKGTAPTTTLKSPAPTTPKKPAKELAPTTTKEPTTTTSDKPAPTTKGTAPTT 687
QY 781 TPKETAPTTKEAPTTPKKPAKELAPTTTKEPTTTTSDKPAPTTKGTAPTTTSDKPAPTTKGTAPTT 840
Db 688 TPKETAPTTKEAPTTPKKPAKELAPTTTKEPTTTTSDKPAPTTKGTAPTTTSDKPAPTTKGTAPTT 747
QY 841 AEPTPKALENSKPEKGVPTTKTAAATKPEMTTAKDKTERDLRTTPETTTAAAPKMTKET 900
Db 748 AEPTPKALENSKPEKGVPTTKTAAATKPEMTTAKDKTERDLRTTPETTTAAAPKMTKET 807
QY 901 ATTTEKTTESKITATTQVTSITTTQDTTTPFKITTLTKTTTLAPKVTITTKTITTTIMNKP 960
Db 808 ATTTEKTTESKITATTQVTSITTTQDTTTPFKITTLTKTTTLAPKVTITTKTITTTIMNKP 867
QY 961 EETAKPKDRATNSKATTPKPKQPTKAPKPTSTKPKTMPRVRKPKTTTPKPKMTSTTPE 1020
Db 868 EETAKPKDRATNSKATTPKPKQPTKAPKPTSTKPKTMPRVRKPKTTTPKPKMTSTTPE 927
QY 1021 LNPTSRIAEAMLOTTTRPNQTNENSKLVEVNPKESEDAGGAEGETPHMLLRPHVMEVEVTPD 1080
Db 928 LNPTSRIAEAMLOTTTRPNQTNENSKLVEVNPKESEDAGGAEGETPHMLLRPHVMEVEVTPD 987
QY 1081 MDYLPVNPQGIINPMLS 1099
Db 988 MDYLPVNPQGIINPMLS 1006

```

Search completed: October 13, 2004, 11:58:45
Job time : 25.1039 secs


```

; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104

Query Match          99.8%; Score 5820.9; DB 13; Length 1140;
Best Local Similarity 96.4%; Pred. No. 5.2e-142;
Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIQVSSQ----- 25
DB 1 MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSDATCNCYDNCQHYMECCPDF 60

QY 26 -----ELSCGRGCFESFERGECDDAQQCKYDKCCPDYEFCAEVHNPTSPSSSKAP 79
DB 61 KRVTAEELSCGRGCFESFERGECDDAQQCKYDKCCPDYEFCAEVHNPTSPSSSKAP 120

QY 80 PPASGASQTIKSTTKSPKPNKKTKVIESEITEHSVSENQESSSSSSSSSTI 139
DB 121 PPASGASQTIKSTTKSPKPNKKTKVIESEITEHSVSENQESSSSSSSSSTI 180

QY 140 KIKSSKNSAANRELQKLVKDKNKNRTKKKPTPKPPVVDAGSLDNGDFKVTPTST 199
DB 181 KIKSSKNSAANRELQKLVKDKNKNRTKKKPTPKPPVVDAGSLDNGDFKVTPTST 240

QY 200 TOHNVKSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKKETTTNKQSTSDG 259
DB 241 TOHNVKSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKKETTTNKQSTSDG 300

QY 260 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPK 319
DB 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPK 360

QY 320 TTPKEPTPTTIKSAPTTPKPAETTTKGAPTTKPSAPTTTKBPAPTTTKBPAPTTTKEP 379
DB 361 TTPKEPTPTTIKSAPTTPKPAETTTKGAPTTKPSAPTTTKBPAPTTTKBPAPTTTKEP 420

QY 380 APPTTKSAPTTTPKBPAPTTTPKPAETTTKGPALTTKPEPTTPKBPAPTTTPK 439
DB 421 APPTTKSAPTTTPKBPAPTTTPKPAETTTKGPALTTKPEPTTPKBPAPTTTPK 480

QY 440 EPAPTAPKPAETTTKBPAPTTTPKPAETTTKGPALTTKPEPTTPKBPAPTTTPK 499
DB 481 EPAPTAPKPAETTTKBPAPTTTPKPAETTTKGPALTTKPEPTTPKBPAPTTTPK 540

QY 500 TTKSAPTTTPKBPAPTTTPKPAETTTKGPALTTKPEPTTPKBPAPTTTPKBPAPTTTPK 559
DB 541 TTKSAPTTTPKBPAPTTTPKPAETTTKGPALTTKPEPTTPKBPAPTTTPKBPAPTTTPK 600

QY 560 APAPTAPKPAETTTKBPAPTTTPKPAETTTKGPALTTKPEPTTPKBPAPTTTPK 619
DB 601 APAPTAPKPAETTTKBPAPTTTPKPAETTTKGPALTTKPEPTTPKBPAPTTTPK 660

QY 620 PEEAPTTTPKAAAPNTPKBPAPTTTPKPAETTTKGPALTTKPEPTTPKBPAPTTTPK 679
DB 661 PEEAPTTTPKAAAPNTPKBPAPTTTPKPAETTTKGPALTTKPEPTTPKBPAPTTTPK 720

QY 680 APPTPKKAPKELAPTTTKPEPTTSDKPAETTTKGPALTTKPEPTTPKBPAPTTTPK 739
DB 721 APPTPKKAPKELAPTTTKPEPTTSDKPAETTTKGPALTTKPEPTTPKBPAPTTTPK 780

QY 740 TAPPTLKAPAPTTTPKPAETTTKGPALTTKPEPTTPKBPAPTTTPKBPAPTTTPK 799
DB 781 TAPPTLKAPAPTTTPKPAETTTKGPALTTKPEPTTPKBPAPTTTPKBPAPTTTPK 840

QY 800 KPAPTTPETPTTSEVSTPTTKPEPTTHKSPDESSTPELSAETTPKALENSPKPGVPT 859
DB 841 KPAPTTPETPTTSEVSTPTTKPEPTTHKSPDESSTPELSAETTPKALENSPKPGVPT 900

QY 860 TKTPTAKTEMTTAKDKTTERDLRTTPTTTAAAPKMTKETATTTKTESKITATTTQV 919
DB 901 TKTPTAKTEMTTAKDKTTERDLRTTPTTTAAAPKMTKETATTTKTESKITATTTQV 960

QY 920 TSTTTQDTPFPKITTLLKTTLLAPKVTTTKKTTTTTTEIMNKPBEATKPKDRATNSKATTPK 979

```

```

DB 961 TSTTTQDTPFPKITTLLKTTLLAPKVTTTKKTTTTTTEIMNKPBEATKPKDRATNSKATTPK 1020
QY 980 PQPFTAPKPKPTSTKPKTMPVRKPKTTTTPRKMSTSTPELNPTSRIAEAMLOTTTRN 1039
DB 1021 PQPFTAPKPKPTSTKPKTMPVRKPKTTTTPRKMSTSTPELNPTSRIAEAMLOTTTRN 1080
QY 1040 QTPNSKLVEVNPKSESDAGGAGETPHMLLRPHVFMPEVTPDMYDLPRVFNQGIINPMLS 1099
DB 1081 QTPNSKLVEVNPKSESDAGGAGETPHMLLRPHVFMPEVTPDMYDLPRVFNQGIINPMLS 1140

RESULT 3
US-09-802-207-30
; Sequence 30, Application US/09802207
; Publication No. US20020086824A1
; GENERAL INFORMATION:
; APPLICANT: Warman, Matthew
; APPLICANT: Carpten, John
; APPLICANT: Trent, Jeffrey
; APPLICANT: Marcelino, Jose
; TITLE OF INVENTION: Novel Methods and Reagents for the Treatment of Osteoarthritis
; FILE REFERENCE: Case-06212
; CURRENT APPLICATION NUMBER: US/09/802,207
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 09/619,175
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/145,328
; PRIOR FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 30
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-802-207-30

Query Match          99.8%; Score 5820.9; DB 9; Length 1404;
Best Local Similarity 96.4%; Pred. No. 6.6e-142;
Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIQVSSQ----- 25
DB 1 MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSDATCNCYDNCQHYMECCPDF 60

QY 26 -----ELSCGRGCFESFERGECDDAQQCKYDKCCPDYEFCAEVHNPTSPSSSKAP 79
DB 61 KRVTAEELSCGRGCFESFERGECDDAQQCKYDKCCPDYEFCAEVHNPTSPSSSKAP 120

QY 80 PPASGASQTIKSTTKSPKPNKKTKVIESEITEHSVSENQESSSSSSSSSTI 139
DB 121 PPASGASQTIKSTTKSPKPNKKTKVIESEITEHSVSENQESSSSSSSSSTI 180

QY 140 KIKSSKNSAANRELQKLVKDKNKNRTKKKPTPKPPVVDAGSLDNGDFKVTPTST 199
DB 181 KIKSSKNSAANRELQKLVKDKNKNRTKKKPTPKPPVVDAGSLDNGDFKVTPTST 240

QY 200 TOHNVKSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKKETTTNKQSTSDG 259
DB 241 TOHNVKSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKKETTTNKQSTSDG 300

QY 260 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPK 319
DB 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPK 360

QY 320 TTPKEPTPTTIKSAPTTPKPAETTTKGAPTTKPSAPTTTKBPAPTTTKBPAPTTTKEP 379
DB 361 TTPKEPTPTTIKSAPTTPKPAETTTKGAPTTKPSAPTTTKBPAPTTTKBPAPTTTKEP 420

QY 380 APPTTKSAPTTTPKBPAPTTTPKPAETTTKGPALTTKPEPTTPKBPAPTTTPK 439
DB 421 APPTTKSAPTTTPKBPAPTTTPKPAETTTKGPALTTKPEPTTPKBPAPTTTPK 480

```


RESULT 5

US-10-124-557-2
 ; Sequence 2, Application US/10124557
 ; Publication No. US20020137894A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Turner, Katherine
 ; Clark, Stephen C.
 ; Jacobs, Kenneth
 ; Hewick, Rodney M.
 ; Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 CambridgePark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/124,557
 FILING DATE: 16-Apr-2002
 CLASSIFICATION: <Unknown>
 APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 23-JUN-1990
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:
 NAME: Cserr, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 876-1170
 TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1404 amino acids
 TYPE: amino acid
 TOPOLOGY: linear

MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-10-124-557-2

Query Match 99.8%; Score 5820.9; DB 13; Length 1404;
 Best Local Similarity 96.4%; Pred. No. 6.6e-142;
 Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;
 QY 1 MAWKLPYLLLLLVFVIQVSSQ----- 25
 DB 1 MAWKLPYLLLLLVFVIQVSSQDLSSCAGRCGEGYSDATCNCYDNCQHYMECCPDF 60
 QY 26 -----ELSCKGRCFESFERGECDCDAQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 79
 DB 61 KRVTAEJLSCKGRCFESFERGECDCDAQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120
 QY 80 PPSGASQTKSTKSPKPNKKTKVIESEITEHSVSNQSSSSSSSSSSSSSTI 139
 DB 121 PPSGASQTKSTKSPKPNKKTKVIESEITEHSVSNQSSSSSSSSSSSSSTI 180
 QY 140 KTKSSNSAANRELQKLVKDNKKTKKTPPPVVDAGSGLDNGDFKVTTPDTST 199

US-10-124-557-2
 Query Match 99.8%; Score 5820.9; DB 13; Length 1404;
 Best Local Similarity 96.4%; Pred. No. 6.6e-142;
 Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;
 QY 1 MAWKLPYLLLLLVFVIQVSSQ----- 25
 DB 1 MAWKLPYLLLLLVFVIQVSSQDLSSCAGRCGEGYSDATCNCYDNCQHYMECCPDF 60
 QY 26 -----ELSCKGRCFESFERGECDCDAQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 79
 DB 61 KRVTAEJLSCKGRCFESFERGECDCDAQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120
 QY 80 PPSGASQTKSTKSPKPNKKTKVIESEITEHSVSNQSSSSSSSSSSSSSTI 139
 DB 121 PPSGASQTKSTKSPKPNKKTKVIESEITEHSVSNQSSSSSSSSSSSSSTI 180
 QY 140 KTKSSNSAANRELQKLVKDNKKTKKTPPPVVDAGSGLDNGDFKVTTPDTST 199

RESULT 6

US-10-124-557-62
 ; Sequence 62, Application US/10124557
 ; Publication No. US20020137894A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Turner, Katherine
 ; Clark, Stephen C.
 ; Jacobs, Kenneth
 ; Hewick, Rodney M.
 ; Gesner, Thomas G.

181 KIKSSNSAANRELQKLVKDNKKTKKTPPPVVDAGSGLDNGDFKVTTPDTST 240
 QY 200 TOHNVKYSTSPKITTAKPINPRSLPNSDTSKETSITVANKETTVETKETTTNKQSTDG 259
 DB 241 TOHNVKYSTSPKITTAKPINPRSLPNSDTSKETSITVANKETTVETKETTTNKQSTDG 300
 QY 260 KEXTTSKETSQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTPKPPTTPPKEPAS 319
 DB 301 KEXTTSKETSQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTPKPPTTPPKEPAS 360
 QY 320 TTPKEPTPTTIKSAPTTPKAPPTTKSAPTTKAPTTKAPTTKAPTTKAPTTKAPTTKAP 379
 DB 361 TTPKEPTPTTIKSAPTTPKAPPTTKSAPTTKAPTTKAPTTKAPTTKAPTTKAPTTKAP 420
 QY 380 APTTTKSAPTTKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAP 439
 DB 421 APTTTKSAPTTKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAP 480
 QY 440 EPAPTAPKKPAPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPK 499
 DB 481 EPAPTAPKKPAPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPK 540
 QY 500 TTKSAPTTPKPSPTTKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPK 559
 DB 541 TTKSAPTTPKPSPTTKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPK 600
 QY 560 APTAPKAPPTPKETAPTTPKLTTPKLTTPKLTTPKLTTPKLTTPKLTTPKLTTPKLT 619
 DB 601 APTAPKAPPTPKETAPTTPKLTTPKLTTPKLTTPKLTTPKLTTPKLTTPKLTTPKLT 660
 QY 620 PEPAPTTPKAAAPNTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPK 679
 DB 661 PEPAPTTPKAAAPNTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPK 720
 QY 680 APTPKKAPKELAPTTPKPTSTTSKAPPTPKGTAPTTPKAPPTTPKAPPTTPKAPPTPK 739
 DB 721 APTPKKAPKELAPTTPKPTSTTSKAPPTPKGTAPTTPKAPPTTPKAPPTTPKAPPTPK 780
 QY 740 TAPTTLKAPPTTPKAPKELAPTTPKPTSTTSKAPPTPKGTAPTTPKAPPTTPKAPPTTPK 799
 DB 781 TAPTTLKAPPTTPKAPKELAPTTPKPTSTTSKAPPTPKGTAPTTPKAPPTTPKAPPTTPK 840
 QY 800 KPAPTTPETPTTSEVSTPTTKPTTIHKSPDESTPELSAPPTPKALENSPKPGVPT 859
 DB 841 KPAPTTPETPTTSEVSTPTTKPTTIHKSPDESTPELSAPPTPKALENSPKPGVPT 900
 QY 860 TKTPAAKPEMTTAKDKITERDLRTPETTTAAKPMKETAATTTESKITATTQV 919
 DB 901 TKTPAAKPEMTTAKDKITERDLRTPETTTAAKPMKETAATTTESKITATTQV 960
 QY 920 TSTTTQDTPFKITTLKTTTLAPKVTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 979
 DB 961 TSTTTQDTPFKITTLKTTTLAPKVTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 1020
 QY 980 PQKPTKAPKPTSTKPKMTMPRVKPKTTPPRKMTSTWPELNPSTSRISAMLOTTTRPN 1039
 DB 1021 PQKPTKAPKPTSTKPKMTMPRVKPKTTPPRKMTSTWPELNPSTSRISAMLOTTTRPN 1080
 QY 1040 QTPNSKLVEYNPKSEDAAGGAGETPHMLLPHVFMPEVTPDMVDYLPVNPQGIINPMLS 1099
 DB 1081 QTPNSKLVEYNPKSEDAAGGAGETPHMLLPHVFMPEVTPDMVDYLPVNPQGIINPMLS 1140

; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 Cambridgepark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ;
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/124,557
 ; FILING DATE: 16-Apr-2002
 ; CLASSIFICATION: <Unknown>
 ;
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; APPLICATION NUMBER: US 07/546,114
 ; FILING DATE: 29-JUN-1990
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cserr, Luann
 ; REGISTRATION NUMBER: 31,822
 ; REFERENCE/DOCKET NUMBER: GI 5190
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)876-1170
 ; TELEFAX: (617)876-5851
 ;
 ; INFORMATION FOR SEQ ID NO: 62:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1404 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 62:
 US-10-124-557-62

Query Match 99.8%; Score 5820.9; DB 13; Length 1404;

Best Local Similarity 96.4%; Pred. No. 6.6e-142;
Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSGQ----- 25
 DB 1 MAWKTLPIYLLLLSVFVIQVSSQDILSSCAGRCGYSRATCNCDCYNQCHYMECCPDF 60
 QY 26 -----ELSCGRCFESPERGECDCDAQCKYDKCCPDYSEFCAEVHNPTSPSSKAP 79
 DB 61 KRVTAEELSCGRCFESPERGECDCDAQCKYDKCCPDYSEFCAEVHNPTSPSSKAP 120
 QY 80 PPSGASQIKSTTKRSPKPNKXKTKVIBSEETEEHSVSENOESSSSSSSSSSSTI 139
 DB 121 PPSGASQIKSTTKRSPKPNKXKTKVIBSEETEEHSVSENOESSSSSSSSSSSTI 180
 QY 140 KIKSKNSAANRELQKLVKXDNKNKTKKPTPKPPVDEAGSLDNGDFKVTTPDTST 199
 DB 181 KIKSKNSAANRELQKLVKXDNKNKTKKPTPKPPVDEAGSLDNGDFKVTTPDTST 240
 QY 200 TQHNKVTSPKITTAKPINRPSLPKNSDTSKETSLSLVNKETTVETKETTITNKQSTDG 259
 DB 241 TQHNKVTSPKITTAKPINRPSLPKNSDTSKETSLSLVNKETTVETKETTITNKQSTDG 300
 QY 260 KEKTTSAKETOSIEKTSKADLAPTS KVLAKPTPKAETTTKGPALTTPKEPTTPKPEPAS 319
 DB 301 KEKTTSAKETOSIEKTSKADLAPTS KVLAKPTPKAETTTKGPALTTPKEPTTPKPEPAS 360
 QY 320 TTPKEPTPTTKSAPTTKKEPAPTTTKSAPTTKKEPAPTTKKEPAPTTKKEPAPTTTKEP 379

DB 361 TTPKEPTPTTKSAPTTKKEPAPTTTKSAPTTKKEPAPTTTKEPAPTTTKEPAPTTTKEP 420
 QY 380 APTTTKSAPTTKKEPAPTTPKKAPPTTTPKEPAPTTPKBPTTTPKEPAPTTTKEPAPTTPK 439
 DB 421 APTTTKSAPTTKKEPAPTTPKKAPPTTTPKEPAPTTTKEPAPTTTTPKEPAPTTTKEPAPTTPK 480
 QY 440 EPAPTAPKAPAPTTTPKEPAPTTPKBAPTTTKEPAPTTTKEPAPTTTTPKEPAPTTTKEPAPTT 499
 DB 481 EPAPTAPKAPAPTTTPKEPAPTTPKBAPTTTKEPAPTTTKEPAPTTTTPKEPAPTTTKEPAPTT 540
 QY 500 TTKSAPTTTPKEPAPTTTTPKEPAPTTTPKAPPTTTPKEPAPTTTPKAPPTTTPKEPAPTTTTPK 559
 DB 541 TTKSAPTTTPKEPAPTTTTPKEPAPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPK 600
 QY 560 APTAPKAPAPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPK 619
 DB 601 APTAPKAPAPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPKAPPTTTPK 660
 QY 620 PEEPAPTTTPKAAAPNTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 679
 DB 661 PEEPAPTTTPKAAAPNTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 720
 QY 680 APTTPKAPKAPKELAPTTTKEPAPTTTSDKAPPTTPKGTAPPTTPKGTAPPTTPKGTAPPTTPKGTAPPTTPK 739
 DB 721 APTTPKAPKAPKELAPTTTKEPAPTTTSDKAPPTTPKGTAPPTTPKGTAPPTTPKGTAPPTTPKGTAPPTTPK 780
 QY 740 TAPTTLKEPAPTTTPKAPKAPKELAPTTTKEPAPTTTSDKAPPTTPKGTAPPTTPKGTAPPTTPKGTAPPTTPK 799
 DB 781 TAPTTLKEPAPTTTPKAPKAPKELAPTTTKEPAPTTTSDKAPPTTPKGTAPPTTPKGTAPPTTPKGTAPPTTPK 840
 QY 800 KAPPTTPPTTPPTTSEVSTPTTTPKPTTIHKSPDSTPSELSPDSTPSELSPDSTPSELSPDSTPSELSPDSTPSEL 859
 DB 841 KAPPTTPPTTPPTTSEVSTPTTTPKPTTIHKSPDSTPSELSPDSTPSELSPDSTPSELSPDSTPSELSPDSTPSEL 900
 QY 860 TKTPAATKEPMTTAKDKTTERDLATTPETTTTAAEPKMTKETATTTTEKTESKITATTTQV 919
 DB 901 TKTPAATKEPMTTAKDKTTERDLATTPETTTTAAEPKMTKETATTTTEKTESKITATTTQV 960
 QY 920 TSTTTQDTPPKITTLKTTTLAPKVTITTKITTTTINMKEPETAAPKXDRATNSKATTPK 979
 DB 961 TSTTTQDTPPKITTLKTTTLAPKVTITTKITTTTINMKEPETAAPKXDRATNSKATTPK 1020
 QY 980 POKPTKAPKPTSTKPKTTPRVRKPTTPPKMTSTMPBLNPTSRFAEAMLOTTTPN 1039
 DB 1021 POKPTKAPKPTSTKPKTTPRVRKPTTPPKMTSTMPBLNPTSRFAEAMLOTTTPN 1080
 QY 1040 QTPNSKLVENVPKSDAGAGETPHMLLRPHVFMPEVTPMDVLPVFNQGIINPMLS 1099
 DB 1081 QTPNSKLVENVPKSDAGAGETPHMLLRPHVFMPEVTPMDVLPVFNQGIINPMLS 1140

RESULT 7

US-10-124-557-46

; Sequence 46, Application US/10124557

; Publication No. US20020137894A1

; GENERAL INFORMATION:

; APPLICANT: Turner, Katherine

; Jacobs, Kenneth

; Hewick, Rodney M.

; Gesner, Thomas G.

; NUMBER OF SEQUENCES: 143

; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 Cambridgepark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02140

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/10/124,557
  FILING DATE: 16-Apr-2002
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/643,502
  FILING DATE: 18-JAN-1991
  APPLICATION NUMBER: US 07/546,114
  FILING DATE: 29-DEC-1989
  APPLICATION NUMBER: US 07/390,901
  FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
  NAME: Cserek, Luann
  REGISTRATION NUMBER: 31,822
  REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (617)876-1170
  TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 46:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1320 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
  MOLECULE TYPE: protein
  SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-124-557-46

Query Match      96.3%; Score 5617.7; DB 13; Length 1320;
Best Local Similarity 96.1%; Pred. No. 1e-136;
Matches 1056; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSCGRCFESFERGRCDCDAQCKYDKCCPDYE 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQELSCGRCFESFERGRCDCDAQCKYDKCCPDYE 60

QY 61 SFCAEVHNPSTPPSKKAPPPSGASQTKSTTKRSPKPPNKKYKVVESBEITEHSVS 120
DB 61 SFCAEVHNPSTPPSKKAPPPSGASQTKSTTKRSPKPPNKKYKVVESBEITE 115

QY 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKDKNKRTKKKTPKPPVYDE 180
DB 116 -----VKUNKNRTKKKTPKPPVYDE 137

QY 181 AGSGLDNGDFKVTPTDSTTOHNVSTSPKITTAKPINRPSLPNSDTSKETSITVYKE 240
DB 138 AGSGLDNGDFKVTPTDSTTOHNVSTSPKITTAKPINRPSLPNSDTSKETSITVYKE 197

QY 241 TTVETKEVTTNKQSTDGKETTSAKETOSIEKTSKADLAPTSKVLAKEPTKAEITTKG 300
DB 198 TTVETKEVTTNKQSTDGKETTSAKETQSIEKTSKADLAPTSKVLAKEPTKAEITTKG 257

QY 301 PALTTPKETPTTPKEPASSTTPKEPTPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTT 360
DB 258 PALTTPKETPTTPKEPASSTTPKEPTPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTT 317

QY 361 KEPAPTTPKEPAPTTTPKEPAPTTTKSAPTTTPKEPAPTTPKKAPAPTTTPKEPAPTTTPKEPTP 420
DB 318 KEPAPTTPKEPAPTTTPKEPAPTTTKSAPTTTPKEPAPTTPKKAPAPTTTPKEPAPTTTPKEPTP 377

QY 421 TTPKEPAPTTTPKEPAPTTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPTPKE 480
DB 378 TTPKEPAPTTTPKEPAPTTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPTPKE 437

QY 481 PAPTTPKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKAPAPTTTP 540
DB 438 PAPTTPKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKAPAPTTTP 497

QY 541 KEPAPTTPKEPAPTTTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 600
DB 541 KEPAPTTPKEPAPTTTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 600
```

RESULT 8

US-10-124-557-60

; Sequence 60, Application US/10124557

; Publication No. US20020137894A1

; GENERAL INFORMATION:

; APPLICANT: Turner, Katherine

; Clark, Stephen C.

; Jacobs, Kenneth

; Hewick, Rodney M.

; Gesner, Thomas G.

; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; FILING DATE: 16-Apr-2002

; APPLICATION NUMBER: US/10/124,557

; CLASSIFICATION: <unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/643,502

; FILING DATE: 18-JAN-1991

; APPLICATION NUMBER: US 07/546,114

; FILING DATE: 29-JUN-1990

; APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 1320 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULAR TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-124-557-60

Query Match
Best Local Similarity 96.3%; Score 5617.7; DB 13; Length 1320;
Matches 1056; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQVSSQELSCKRCFESPERGREGDCDAQCKKYDKCCPDYE 60
DB 1 MAWKTLPIYLLLSVFIQVSSQELSCKRCFESPERGREGDCDAQCKKYDKCCPDYE 60
QY 61 SFCAEVHNPSPSSSKAPPSPGASQTIKSTKSPKPNKKTKKVIESEITEHSVS 120
DB 61 SFCAEVHNPSPSSSKAPPSPGASQTIKSTKSPKPNKKTKKVIESEITE----- 115
QY 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELOKLVKDKNKKNRKTKKPTPPPVVDE 180
DB 116 -----VKDNKKNRKTKKPTPPPVVDE 137
QY 181 AGSGLDNGDFKVTTPDSTTOHNVKSTSPKITTAKPINRPSLPNSDTSKETSITVKE 240
DB 138 AGSGLDNGDFKVTTPDSTTOHNVKSTSPKITTAKPINRPSLPNSDTSKETSITVKE 197
QY 241 TVTETKETTINKOTSDGKEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTG 300
DB 198 TVTETKETTINKOTSDGKEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTG 257
QY 301 PALTPKEPTTPKPEASTTPKEPTPTTIKSAPTTPKEPAPTTKSAPTTPKEPAPTTT 360
DB 258 PALTPKEPTTPKPEASTTPKEPTPTTIKSAPTTPKEPAPTTKSAPTTPKEPAPTTT 317
QY 361 KEPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 420
DB 318 KEPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 377
QY 421 TTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 480
DB 378 TTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 437
QY 481 PAPTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 540
DB 438 PAPTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 497
QY 541 KEPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 600
DB 498 KEPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 557
QY 601 PTTPEELAPTTPEPTTPPEPAPTTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 660
DB 558 PTTPEELAPTTPEPTTPPEPAPTTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 617
QY 661 KETAPTTPKGTAPTTLKEPAPTTTPKKAPKELAPTTTPKSTTSKAPTTTPKGTAPTT 720
DB 618 KETAPTTPKGTAPTTLKEPAPTTTPKKAPKELAPTTTPKSTTSKAPTTTPKGTAPTT 677
QY 721 PKEPAPTTTPKEPAPTTTPKGTAPTTLKEPAPTTTPKKAPKELAPTTTPKSTTSKAPTT 780

678 PKEPAPTTTPKEPAPTTTPKGTAPTTLKEPAPTTTPKKAPKELAPTTTPKSTTSKAPTT 737
781 TPKETAPTTTPKSPAPTTTPKKAPAPTTTPETPTTTPSEVSTTTTKEPTTIHKSPDESTPELS 840
738 TPKETAPTTTPKSPAPTTTPKKAPAPTTTPETPTTTPSEVSTTTTKEPTTIHKSPDESTPELS 797
841 AEPTPKALENSPKPCVPPTTKPAATKPEMTTAKDITTERDLRTTPETTTAAPKMTKET 900
798 AEPTPKALENSPKPCVPPTTKPAATKPEMTTAKDITTERDLRTTPETTTAAPKMTKET 857
901 ATTETKTESKITATTQVTSITTTQDTPFKITLTKITTLAPKVTTKKTIITTEIMNKP 960
858 ATTETKTESKITATTQVTSITTTQDTPFKITLTKITTLAPKVTTKKTIITTEIMNKP 917
961 EETAKPKDRATNSKATTPKPKQPTKAPKPTSTKPKKPNRVRKPKXTTPTPKMTSTMPE 1020
918 EETAKPKDRATNSKATTPKPKQPTKAPKPTSTKPKKPNRVRKPKXTTPTPKMTSTMPE 977
1021 LNPTSRIAEAMLOTTTRPNQNTENSKLVEVNPKSEDAGGAGETPHMLLRPHVMEVETPD 1080
978 LNPTSRIAEAMLOTTTRPNQNTENSKLVEVNPKSEDAGGAGETPHMLLRPHVMEVETPD 1037
1081 MDYLPVPVNPQGIINPMLS 1099
1038 MDYLPVPVNPQGIINPMLS 1056

RESULT 9
US-10-124-557-40
; Sequence 40, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:


```

;      LENGTH: 1361 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-124-557-40

Query Match
Best Local Similarity 96.0%; Score 5603.6; DB 13; Length 1361;
Matches 1056; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYRDATCNDCYNCHYMECCPDF 60

QY 26 -----ELSKGRCFESPERGECDCDAQCKYDKCCDPDYESFCAEVNPTSPSSKKAP 79
DB 61 KSVTAELSKGRCFESPERGECDCDAQCKYDKCCDPDYESFCAEVNPTSPSSKKAP 120

QY 80 PPSGASQIKSTTKSPKPNKKTKVIESEITEEHSVSENQSSSSSSSSSSSSSIW 139
DB 121 PPSGASQIKSTTKSPKPNKKTKVIESEITE----- 156

QY 140 KIKSKNSAANRELOKKLVKDNKNRKKKPTPKPPVVDVDSAGSLDNGDKVTPDTST 199
DB 157 -----VKDNKNRKKKPTPKPPVVDVDSAGSLDNGDKVTPDTST 197

QY 200 TQHNKVSPTKITTAKPINRSLPNSDTSKETSITVNKETTVETKETTINKOTSDG 259
DB 198 TQHNKVSPTKITTAKPINRSLPNSDTSKETSITVNKETTVETKETTINKOTSDG 257

QY 260 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTKGPAITPKPEPTTKPEPAS 319
DB 258 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTKGPAITPKPEPTTKPEPAS 317

QY 320 TTPKEPTPTIISAPTTKEPAPTTTKSAPTTKEPAPTTTKPEAPTTKEPAPTTKEP 379
DB 318 TTPKEPTPTIISAPTTKEPAPTTTKSAPTTKEPAPTTTKPEAPTTKEPAPTTKEP 377

QY 380 APTTKSAPTTKEPAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTKEPAPT 439
DB 378 APTTKSAPTTKEPAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTKEPAPT 437

QY 440 EPAPTKAPKAPTTKEPAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTKEPAPT 499
DB 438 EPAPTKAPKAPTTKEPAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTKEPAPT 497

QY 500 TTKSAPTTKEPSPTTKPEAPTTTKPEAPTTTKKAPTTKEPAPTTTKPEAPTTTKKP 559
DB 498 TTKSAPTTKEPSPTTKPEAPTTTKPEAPTTTKKAPTTKEPAPTTTKPEAPTTTKKP 557

QY 560 APTAPKEPAPTTKETAPTTPKLPTTPEKLAFTTPEKLAFTTPEELAPTTPEEPPTT 619
DB 558 APTAPKEPAPTTKETAPTTPKLPTTPEKLAFTTPEKLAFTTPEELAPTTPEEPPTT 617

QY 620 PREPAPTTKAAAPNTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAP 679
DB 618 PREPAPTTKAAAPNTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAP 677

QY 680 APTTKKAPKELAPTTTKETPTSTSDKAPTTTKGTAFTTPEKAPTTPEKAPTTPKG 739
DB 678 APTTKKAPKELAPTTTKETPTSTSDKAPTTTKGTAFTTPEKAPTTPEKAPTTPKG 737

QY 740 TAPTTKEPAPTTKAPKAPKELAPTTTKGPTSTSDKAPTTTKGTAFTTPEKAPTTPK 799
DB 738 TAPTTKEPAPTTKAPKAPKELAPTTTKGPTSTSDKAPTTTKGTAFTTPEKAPTTPK 797

QY 800 KPAPTTPEPTPTTSVSTPTTKETPTTIHKSPDSTPELSAETPKALENSPKPGVPT 859
DB 798 KPAPTTPEPTPTTSVSTPTTKETPTTIHKSPDSTPELSAETPKALENSPKPGVPT 857

QY 860 TKTPAAKPEMTTAKDKTTERDLRTPETTTAAPKQTKETATTTTEKTESKITATTQV 917
DB 858 TKTPAAKPEMTTAKDKTTERDLRTPETTTAAPKQTKETATTTTEKTESKITATTQV 917
DB 920 TSTTTQDTPPKITTLTKTTLAPKVTTKKTIITTEIMKPEETAKPKORATNSKATTPK 979
DB 918 TSTTTQDTPPKITTLTKTTLAPKVTTKKTIITTEIMKPEETAKPKORATNSKATTPK 977
QY 980 POKPTKAPKPTSTKKPKMVRVRKPTTTPRKMSTMPKMTSTMPKMTSTMPKMTSTMPKMT 1039
DB 978 POKPTKAPKPTSTKKPKMVRVRKPTTTPRKMSTMPKMTSTMPKMTSTMPKMTSTMPKMT 1037
QY 1040 QTNPSKLVNPKSDEAGAGETPHMLRPHVFMDEVTPDMMDYLPRVNPQGIINPMLS 1099
DB 1038 QTNPSKLVNPKSDEAGAGETPHMLRPHVFMDEVTPDMMDYLPRVNPQGIINPMLS 1097

RESULT 10
US-10-124-557-58
; Sequence 58, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
;               Clark, Stephen C.
;               Jacobs, Kenneth M.
;               Gesner, Rodney M.
;               Hewick, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1049 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58

Query Match
Best Local Similarity 95.3%; Score 5561; DB 13; Length 1049;
Matches 1049; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSCGRCFESPERGECDCDAQCKYDKCCPDFE 60
```

Db 1 MAWKTLPIYLLLLSVFVIQQVSSQELSCKGRCFESFERGECDCDAQCKYDKCCPDYE 60
QY 61 SCAEVHNTSPSSKAPPPSGASQTKSTTKRSPKPNKKTKKVIIEEITEHSVS 120
Db 61 SCA-----EHSVS 70
QY 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKONKXRTKKKTPKPPVVD 180
Db 71 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKONKXRTKKKTPKPPVVD 130
QY 181 AGSGLDNGDFKVTTPDTSTQHNKYSTPKITAKPINRPSLPNSDTSKETSILT 240
Db 131 AGSGLDNGDFKVTTPDTSTQHNKYSTPKITAKPINRPSLPNSDTSKETSILT 190
QY 241 TTIVETKETTINKQSTDGKEKTSKAKETOSTEKSADLAPTSKVLAKPTPKAETTTG 300
Db 191 TTIVETKETTINKQSTDGKEKTSKAKETOSTEKSADLAPTSKVLAKPTPKAETTTG 250
QY 301 PALTTPKPTTTPKEPASTTTPKEPTPTTIKSAPTTTPKEPAPTTPKSPAPT 360
Db 251 PALTTPKPTTTPKEPASTTTPKEPTPTTIKSAPTTTPKEPAPTTPKSPAPT 310
QY 361 KEPAATTPKEPAPTTPKSPAPTTPKSAPTTTPKEPAPTTPKSPAPTTPKSP 420
Db 311 KEPAATTPKEPAPTTPKSPAPTTPKSAPTTTPKEPAPTTPKSPAPTTPKSP 370
QY 421 TTPKEPAPTTPKEPAPTTPKSPAPTTPKSAPTTTPKEPAPTTPKSPAPTTP 480
Db 371 TTPKEPAPTTPKEPAPTTPKSAPTTTPKEPAPTTPKSAPTTTPKEPAPT 430
QY 481 PAPTTKSAPTTTPKEPAPTTPKSAPTTTPKSPAPTTPKSPAPTTPKSPAPT 540
Db 431 PAPTTKSAPTTTPKEPAPTTPKSAPTTTPKSPAPTTPKSPAPTTPKSPAPT 490
QY 541 KEPAATTPKEPAPTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 600
Db 491 KEPAATTPKEPAPTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPK 550
QY 601 PTPBELAPTTPKEPAPTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTT 660
Db 551 PTPBELAPTTPKEPAPTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTT 610
QY 661 KETAPTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTP 720
Db 611 KETAPTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTP 670
QY 721 PKEPAPTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTP 780
Db 671 PKEPAPTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTP 730
QY 781 TPKEAPTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTP 840
Db 731 TPKEAPTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPTTTP 790
QY 841 AEPTPKALENSPKPGVFTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPT 900
Db 791 AEPTPKALENSPKPGVFTTTPKSAPTTTPKSAPTTTPKSAPTTTPKSAPT 850
QY 901 ATTTEKTTESKITATTQVTSITQDTTPPKITLTKTLAPKVTITTKITTEIN 960
Db 851 ATTTEKTTESKITATTQVTSITQDTTPPKITLTKTLAPKVTITTKITTEIN 910
QY 961 BETAKPKDRATNSKATTPKOKPKAPKPKTSKPKTTPKPKTTPKPKTTPK 1020
Db 911 BETAKPKDRATNSKATTPKOKPKAPKPKTSKPKTTPKPKTTPKPKTTP 970
QY 1021 LNPTSRIAEAMLOTTTRNQTNPNSKLVENPKSDEGAGBETPHMLLRPHVFM 1080
Db 971 LNPTSRIAEAMLOTTTRNQTNPNSKLVENPKSDEGAGBETPHMLLRPHVFM 1030
QY 1081 MDYLPVRVNOGIINPMLS 1099

Db 1031 MDYLPVRVNOGIINPMLS 1049
RESULT 11
US-10-124-557-142
; Sequence 142, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 142:
US-10-124-557-142
Query Match 95.3%; Score 5561; DB 13; Length 1313;
Best Local Similarity 95.5%; Pred No. 2.9e-135;
Matches 1049; Conservative 0; Mismatches 0; Indels 50; Gaps 1;
QY 1 MAWKTLPIYLLLLSVFVIQQVSSQELSCKGRCFESFERGECDCDAQCKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQELSCKGRCFESFERGECDCDAQCKYDKCCPDYE 60
QY 61 SCAEVHNTSPSSKAPPPSGASQTKSTTKRSPKPNKKTKKVIIEEITEHSVS 120
Db 61 SCA-----EHSVS 70
QY 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKONKXRTKKKTPKPPVVD 180
Db 71 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKONKXRTKKKTPKPPVVD 130
QY 181 AGSGLDNGDFKVTTPDTSTQHNKYSTPKITAKPINRPSLPNSDTSKETSILT 240

```

1  Gesner, Thomas G.
2  TITLE OF INVENTION: Megakaryocyte Stimulating Factors
3  NUMBER OF SEQUENCES: 143
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE: Genetics Institute, Inc.
6  STREET: 87 CambridgePark Drive
7  CITY: Cambridge
8  STATE: Massachusetts
9  COUNTRY: U.S.A.
10 ZIP: 02140
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: PatentIn Release #1.0, Version #1.25
17
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/10/124,557
20 FILING DATE: 16-Apr-2002
21 CLASSIFICATION: <Unknown>
22
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: US 07/643,502
25 FILING DATE: 18-JAN-1991
26 APPLICATION NUMBER: US 07/546,114
27 FILING DATE: 29-JUN-1990
28 APPLICATION NUMBER: US 07/457,196
29 FILING DATE: 29-DEC-1989
30 APPLICATION NUMBER: US 07/390,901
31 FILING DATE: 08-AUG-1989
32
33 ATTORNEY/AGENT INFORMATION:
34 NAME: Cserr, Luann
35 REGISTRATION NUMBER: 31,822
36 REFERENCE/DOCKET NUMBER: GI 5190
37
38 TELECOMMUNICATION INFORMATION:
39 TELEPHONE: (617) 876-1170
40 TELEFAX: (617) 876-5851
41
42 INFORMATION FOR SEQ ID NO: 48:
43 SEQUENCE CHARACTERISTICS:
44 LENGTH: 1354 amino acids
45 TYPE: amino acid
46 TOPOLOGY: linear
47
48 MOLECULE TYPE: protein
49 SEQUENCE DESCRIPTION: SEQ ID NO: 48:
50 PS-10-124-557-48

```

Query Match 95.1%; Score 5546.9; DB 13; Length 1354;
Best Local Similarity 92.0%; Pred. No. 7e-135;
Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

Qy	1	MAWKTLPIYLLLLLVFIQVSSQ-----	25
Db	1	MAWKTLPIYLLLLLVFIQVSSQDLSCAGRCGEGYSRDATCNDYNCQHMECCDPF	60
Qy	26	-----ELSKGRCFESFERGECDDAQKKYDKCCPDYEFCAEVHNPTSPPPSKXAP	79
Db	61	KRVCTAELSKGRCFESFERGECDDAQKKYDKCCPDYEFCA-----	105
Qy	80	PPSGASQIKSTTKESPKNKKTKVIESEIITEHVSSENQESSSSSSSSSSSSSTIW	139
Db	106	-----EHSVSENQESSSSSSSSSSSSSSSTIW	130
Qy	140	KIKSSKNSAANRELOKLKVONKNRTKKTPKPPVVDEAGSLDNGDFKVTTDPDTST	199
Db	131	KIKSSKNSAANRELOKLKVONKNRTKKTPKPPVVDEAGSLDNGDFKVTTDPDTST	130
Qy	200	TOHNKVSTSPKITTAKPINPRSLPNSDTSKETSITVNKEITTVETKEITTTINKQNSTDG	259
Db	191	TOHNKVSTSPKITTAKPINPRSLPNSDTSKETSITVNKEITTVETKEITTTINKQNSTDG	250
Qy	260	KEKTTSAKETQSIKETSANDLAPTQKVLAKPPTKAEITTTKGALTTTPKPEPTPTTPKEPAS	319
Db	251	KEKTTSAKETQSIKETSANDLAPTQKVLAKPPTKAEITTTKGALTTTPKPEPTPTTPKEPAS	310
Qy	320	TTPKPEPTPTTIKSAPTTPKEPAPTITTKSAPTTPKEPAPTITTKPEPAPTTPKEPAPTITTKPE	379

RESULT 12
US-10-124-557-48
; Sequence 48, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.

[illegible]

RESULT 13

US-10-124-557-50

US-10-124-337-30
: Sequence 50. Application US/10124557

; sequence 30, Application US/101
: Publication No. US20020137894A1

; PUBLICATION NO. USZ002
: GENERAL INFORMATION:

GENERAL INFORMATION: APPLICANT: TURNER

APPLICANT: JUTH
Class:

Clar: 1

Jaco.

; Hewi

Gesn.

TITLE OF INVENTION:

NUMBER OF SEQUENCES

CORRESPONDENCE

CORRESPONDENCE: ADDRESS:

ADDRESS: STAFF. 97

STREET: 87
CITY: Camb

; CITY: Camb.

; STATE: Mas

; COUNTRY: U

ZIP: 02140

491 PKPAPPTPKPAPPTTKKPAAPKAPAPPTPKETAPPTPKLTPTTPEKAPPTPKP 550
600 APPTPEELAPPTPEPTTPEEPAPPTPKAAAPNTKPEAPPTPKPAPPTPKPAPPT 659
551 APPTPEELAPPTPEPTTPEEPAPPTPKAAAPNTKPEAPPTPKPAPPTPKPAPPT 610
660 PKETAPPTPKGAPPTTLKPEAPPTPKKAPKELAPPTTKETSTTSKAPPTPKGTAPT 719
611 PKETAPPTPKGAPPTTLKPEAPPTPKKAPKELAPPTTKETSTTSKAPPTPKGTAPT 670
720 TPKEAPPTPKPAPPTPKGAPPTTLKPEAPPTPKKAPKELAPPTTKGPTSTTSKAP 779
671 TPKEAPPTPKPAPPTPKGAPPTTLKPEAPPTPKKAPKELAPPTTKGPTSTTSKAP 730
780 TTPKETAPPTPKPAPPTPKPAPPTPKPPTTSEVSTPTTKEPTTIHKSPDESTPEL 839
731 TTPKETAPPTPKPAPPTPKPAPPTPKPPTTSEVSTPTTKEPTTIHKSPDESTPEL 790
840 SAEPPTPKALENSPKPFGVPTTKPAATKPEMTTAKOKITTERDLTTPETTTAAPKMTKE 899
791 SAEPPTPKALENSPKPFGVPTTKPAATKPEMTTAKOKITTERDLTTPETTTAAPKMTKE 850
900 TATTTEKTESKITATTQVTSITTTQDTTPFKITTLKTTLAPKVTTTKITTEIMNK 959
851 TATTTEKTESKITATTQVTSITTTQDTTPFKITTLKTTLAPKVTTTKITTEIMNK 910
960 PEETAKPDRAATNSKATPKPKPTKAPKPTSTKPKTMRVRKPTTTPRKMSTMP 1019
911 PEETAKPDRAATNSKATPKPKPTKAPKPTSTKPKTMRVRKPTTTPRKMSTMP 970
1020 ELNPTSRRAEAMLTQTTTRNQTNPNSKLVEVNPKSDAGAGETPHMLLRPHVFMPEVTP 1079
971 ELNPTSRRAEAMLTQTTTRNQTNPNSKLVEVNPKSDAGAGETPHMLLRPHVFMPEVTP 1030
1080 DMDYLPVFNQGIINPMLS 1099
1031 DMDYLPVFNQGIINPMLS 1050

RESULT 14

US-10-124-557-74
; Sequence 74, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-10-124-557-74

Query Match

Best Local Similarity 91.8%; Score 5353.7; DB 13; Length 1038;
Matches 100%; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
Qy 1 MANKTLPIYLLLLSVFVIOQVSSQBLSCGRGCFESFERGRCDDAQQCKYDKCCPDYE 60
Db 1 MANKTLPIYLLLLSVFVIOQVSSQBLSCGRGCFESFERGRCDDAQQCKYDKCCPDYE 60
Qy 61 SFCAEVHNPTSPFSSKXAPPPGASQTIKSTTKRSPKPNKKTKKVIIEEIEBHSVS 120
Db 61 SFCAE ----- 65
Qy 121 ENQESSSSSSSSSSSIWKIKSSKNSAANRELQKXKDKNKKNTKKKTPKPPVVDZ 180
Db 66 -----VKDNKKNTKKKTPKPPVVDZ 87
Qy 181 AGSGLDNGDFKVTTPDTSTTQHNVSTSPKITTAKPINRPSLPNNSDTSKETSLSLVNKE 240
Db 88 AGSGLDNGDFKVTTPDTSTTQHNVSTSPKITTAKPINRPSLPNNSDTSKETSLSLVNKE 147
Qy 241 TVETKETTNTKQSTSDGKEKTSIAKQSIKTSIAKOLAPTSKVLAKTPKAEITTKG 300
Db 148 TVETKETTNTKQSTSDGKEKTSIAKQSIKTSIAKOLAPTSKVLAKTPKAEITTKG 207
Qy 301 PALTTKPEPTTPKPEPASTTPKEPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTT 360
Db 208 PALTTKPEPTTPKPEPASTTPKEPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTT 267
Qy 361 KEPAPTTKPEAPTTTKPEAPTTTKSAPTTKPEAPTTTKKAPPTPKPEAPTTPKPEPTP 420
Db 268 KEPAPTTKPEAPTTTKPEAPTTTKSAPTTKPEAPTTTKKAPPTPKPEAPTTPKPEPTP 327
Qy 421 TTPKEAPTTKPEAPTTKPEAPTAAPKAPPTPKPEAPTTKPEAPTTTKKESPTTKE 480
Db 328 TTPKEAPTTKPEAPTTKPEAPTAAPKAPPTPKPEAPTTKPEAPTTTKKESPTTKE 387
Qy 481 PAPTTKSAPTTKPEAPTTTKSAPTTKPEPSPTTKPEAPTTTKPEAPTTTKKAPPTP 540
Db 388 PAPTTKSAPTTKPEAPTTTKSAPTTKPEPSPTTKPEAPTTTKKAPPTPKPEAPTTP 447
Qy 541 KEPAPTTKPEAPTTTKKAPPTPKPEAPTTTKETAPTTPKLTPTTPEKLAETPEKPA 600
Db 448 KEPAPTTKPEAPTTTKKAPPTPKPEAPTTTKETAPTTPKLTPTTPEKLAETPEKPA 507
Qy 601 PTTPEELAPTTPEPTTPEEPAPTTPKAAAPNTKPEAPTTTKPEAPTTTKPEAPTTT 660
Db 508 PTTPEELAPTTPEPTTPEEPAPTTPKAAAPNTKPEAPTTTKPEAPTTTKPEAPTTT 567
Qy 661 KETAPTTKGTAPTTLKPEAPTTPKKAPKELAPTTTKETSTTSKAPPTPKGTAPT 720
Db 568 KETAPTTKGTAPTTLKPEAPTTPKKAPKELAPTTTKETSTTSKAPPTPKGTAPT 627
Qy 721 PKPEAPTTPKPEAPTTTKGTAPTTLKPEAPTTPKKAPKELAPTTTKGTSTTSKAPPT 780

Db 628 PKPAPPTPKPAPPTPKGAPTTTLKEPAPPTPKKAPKELAPTTTKGPTSTTSKAPPT 687
Qy 781 TPKEATPTPKPAPPTPKKAPPTTPPTTSEVSTPTTKGPTTIHKSPDSTPELS 840
Db 688 TPKEATPTPKPAPPTPKKAPPTTPPTTSEVSTPTTKGPTTIHKSPDSTPELS 747
Qy 841 APTPKALENSPKGCVPTTKTAPAKKEMTTAKKTTERRDLRTTPTTAAKPKMTET 900
Db 748 APTPKALENSPKGCVPTTKTAPAKKEMTTAKKTTERRDLRTTPTTAAKPKMTET 807
Qy 901 ATTTEKTTESKITATTQVTSSTTQDTPPFKITTLLKTTTLAPKVTITTKTTTTEIMNKP 960
Db 808 ATTTEKTTESKITATTQVTSSTTQDTPPFKITTLLKTTTLAPKVTITTKTTTTEIMNKP 867
Qy 961 BETAKPKDRATNSKATTPKPKOKPTKAPKPTSTKPKTTPMVRKPKTTPTRKVTSTMP 1020
Db 868 BETAKPKDRATNSKATTPKPKOKPTKAPKPTSTKPKTTPMVRKPKTTPTRKVTSTMP 927
Qy 1021 LNPTSRISAEMLQTTTRPNQTPNSKLVENPKSEDAGGAGETPHMLLRPHVFPVETPD 1080
Db 928 LNPTSRISAEMLQTTTRPNQTPNSKLVENPKSEDAGGAGETPHMLLRPHVFPVETPD 987
Qy 1081 MDYLPRVNPQGIINPMLS 1099
Db 988 MDYLPRVNPQGIINPMLS 1006

RESULT 15
US-10-124-557-44
; Sequence 44, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserit, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 44:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1270 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-124-557-44

Query Match 91.8%; Score 5353.7; DB 13; Length 1270;
Best Local Similarity 91.5%; Pred. No. 66-130;
Matches 1006; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

Qy 1 MAWKTLPIYLLLLSVFVIQVSSOELSCCKGRCFBSFGRGECDCDAQCKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLLSVFVIQVSSOELSCCKGRCFBSFGRGECDCDAQCKYDKCCPDYE 60
Qy 61 SPCAEVHNFTSPSSKKAPPPSGASQTIKSTTKRSKPKPKKKTKKVIIESEITEHSVS 120
Db 61 SPCAE----- 65
Qy 121 ENQESSSSSSSSSSSTIWKIKSSKNSAANRELQKKLVKONKKKRTKKKTPKPPVDE 180
Db 66 -----VKONKKKRTKKKTPKPPVDE 87
Qy 181 AGSGLDNGDFKVTTPDTSTTOHNVKSTSPKITTAKPINRPSLPNSDTSKETSITVNKE 240
Db 88 AGSGLDNGDFKVTTPDTSTTOHNVKSTSPKITTAKPINRPSLPNSDTSKETSITVNKE 147
Qy 241 TTVEKTEITTKQTSQDCKEKTTSKETSQTSIEKTSADLAPTSKVLAKPTPKAETTKG 300
Db 148 TTVEKTEITTKQTSQDCKEKTTSKETSQTSIEKTSADLAPTSKVLAKPTPKAETTKG 207
Qy 301 PALTTPKKEPTPTTPKEPASTTTPKEPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTT 360
Db 208 PALTTPKKEPTPTTPKEPASTTTPKEPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTT 267
Qy 361 KSPAPPTPKPAPPTTKBPAPTTTKSAPTTTPKEPAPTTPKPAPTTPKPAPTTPKPAPTT 420
Db 268 KSPAPPTPKPAPTTTKBPAPTTTKSAPTTTPKEPAPTTPKPAPTTPKPAPTTPKPAPTT 327
Qy 421 TTPKEPAPTTTKBPAPTTPKPAPTTAPKAPPTTPKEPAPTTPKPAPTTPKPAPTTTPKE 480
Db 328 TTPKEPAPTTTKBPAPTTPKPAPTTAPKAPPTTPKEPAPTTPKPAPTTPKPAPTTTPKE 387
Qy 481 PAPTTPKSAPTTTKBPAPTTTKSAPTTTPKEPAPTTPKPAPTTPKPAPTTPKPAPTTTP 540
Db 388 PAPTTPKSAPTTTKBPAPTTTKSAPTTTPKEPAPTTPKPAPTTPKPAPTTPKPAPTTTP 447
Qy 541 KSPAPTTPKPAPTTTKBPAPTTAPKAPPTTPKEPAPTTPKPAPTTPKPAPTTTPKEPA 600
Db 448 KSPAPTTPKPAPTTTKBPAPTTAPKAPPTTPKEPAPTTPKPAPTTPKPAPTTTPKEPA 507
Qy 601 PTTPEELAPTTPEEPTPTTPPEPAPTTTPKAAAPNTPKPAPTTPKPAPTTTPKEPAPTT 660
Db 508 PTTPEELAPTTPEEPTPTTPPEPAPTTTPKAAAPNTPKPAPTTPKPAPTTTPKEPAPTT 567
Qy 661 KETAPTTKGTAPTTTKBPAPTTPKPAPKELAPTTTKEPTSTTSKAPTPKGTAPT 720
Db 568 KETAPTTKGTAPTTTKBPAPTTPKPAPKELAPTTTKEPTSTTSKAPTPKGTAPT 627
Qy 721 PKPAPTTPKPAPTTTPKGTAPTTLKEPAPTTTPKAPKELAPTTTKGPTSTTSKAPPT 780
Db 628 PKPAPTTPKPAPTTTPKGTAPTTLKEPAPTTTPKAPKELAPTTTKGPTSTTSKAPPT 687
Qy 781 TPKEATPTPKPAPPTPKKAPPTTPPTTSEVSTPTTKGPTTIHKSPDSTPELS 840
Db 688 TPKEATPTPKPAPPTPKKAPPTTPPTTSEVSTPTTKGPTTIHKSPDSTPELS 747
Qy 841 APTPKALENSPKGCVPTTKTAPAKKEMTTAKKTTERRDLRTTPTTAAKPKMTET 900
Db 748 APTPKALENSPKGCVPTTKTAPAKKEMTTAKKTTERRDLRTTPTTAAKPKMTET 807
Qy 901 ATTTEKTTESKITATTQVTSSTTQDTPPFKITTLLKTTTLAPKVTITTKTTTTEIMNKP 960

Db	808	ATTEKTSK	TATTTQVSTTTQD	TTTTPFKITL	TKTTLAPK	VTTTKITITTTIMNK	P	867
QY	961	ETAKPKDRATNSK	ATTPKPKPTKAPK	KPTSTKKPK	TMPRVRK	PKTTTPPK	AKTSTWPE	1020
Db	868	EETAKPKDRATNSK	ATTPKPKPTKAPK	KPTSTKKPK	TMPRVRK	PKTTTPPK	AKTSTWPE	927
QY	1021	LNPSTSIARAMLQ	TTTRPNQ	TNSKLVE	NPKSEDAGG	AGETPHML	LPHVFMPEVTPD	1080
Db	928	LNPSTSIARAMLQ	TTTRPNQ	TNSKLVE	NPKSEDAGG	AGETPHML	LPHVFMPEVTPD	987
QY	1081	MDYLPV	PNQGIINP	MLS			1099	
Db	988	MDYLPV	PNQGIINP	MLS			1006	

Search completed: October 13, 2004, 11:52:38
Job time : 117.81 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 22.3855 Seconds
(without alignments)
4723.689 Million cell updates/sec

Title: SEQ1-A
Perfect score: 5835
Sequence: 1 NAWKTLPIYLLLSVFIQ.....DMDLPRVPGIINPMLS 1099

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1283.7	22.0	3020	A43932	mucin 2 precursor,
2	1143.3	19.6	1664	T18262	S-layer protein -
3	1015.5	17.4	1274	T16251	hypothetical prote
4	1014.9	17.4	2187	T30826	nascent polypeptid
5	1009.9	17.3	1489	T31108	cyst germination s
6	1008.4	17.3	7862	T38346	elastic titin - hu
7	994.8	17.0	3570	T45025	mucin MUC5B, trach
8	933.8	16.0	1367	S48478	glucan 1,4-alpha-g
9	896.2	15.4	3507	T34513	hypothetical prote
10	886.7	15.2	1188	S49915	extensin-like prot
11	879.9	15.1	6642	T23757	protein UNC-89 - C
12	875.8	15.0	2897	T48666	cell proliferation
13	871.4	14.9	1229	T25697	hypothetical prote
14	863.5	14.8	3256	A48666	cell proliferation
15	860.7	14.8	1151	T18535	high molecular mas
16	848.1	14.5	5762	A41819	proline-rich pepti
17	837.8	14.4	1344	A35175	mucin 1 precursor,
18	794.4	13.6	4135	T23629	tenascin-X - bovin
19	794.1	13.6	2232	T34434	hypothetical prote
20	792.1	13.6	2142	E35098	MHC class III hist
21	785.9	13.5	1832	T31113	mucin-like glycop
22	777.7	13.3	990	IS1618	nucleolar phosphop
23	775.7	13.3	3942	T42730	Basoon protein -
24	773.2	13.3	2774	A43359	microtubule-associ
25	766	13.1	1872	S36152	MHC class III hist
26	760.2	13.0	1870	S37671	MHC class III hist
27	756.1	13.0	761	C84672	hypothetical prote
28	751.6	12.9	4548	S00657	apoptobin(a) (EC
29	748.4	12.8	5362	T03454	ALR protein - huma

30	745.7	12.8	3938	2	T42761	Basoon protein -
31	745.1	12.8	2225	2	T26063	hypothetical prote
32	741.5	12.7	924	2	S27923	gene IF3 protein -
33	741.5	12.7	971	2	T19431	hypothetical prote
34	737.4	12.6	1777	2	T34369	hypothetical prote
35	731.6	12.5	1611	2	T38236	hypothetical prote
36	725.2	12.4	5105	2	T32650	hypothetical prote
37	723.9	12.4	4957	2	T03455	ALR protein - huma
38	722.7	12.4	1630	2	A53577	ascites sialoglyco
39	720.8	12.4	3381	2	T42389	versican precursor
40	718.6	12.3	5170	2	T15348	hypothetical prote
41	718	12.3	801	2	T29018	hypothetical prote
42	716.9	12.3	4006	2	T09070	probable tenascin
43	716.8	12.3	3869	2	A48205	All-1 protein +GTE
44	709.7	12.2	3968	2	A44265	trichorax homolog
45	705.8	12.1	1952	2	T48814	hypothetical prote

ALIGNMENTS

RESULT 1

A43932 mucin 2 precursor, intestinal - human (fragments)

N;Alternate names: mucin SMUC-41

C;Species: Homo sapiens (man)

C;Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 09-Jul-2004

C;Accession: A49963; A45106; B45106; A43932; B33532; A61257; PQ0328; PQ0329

R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.

J. Biol. Chem. 269, 2440-2446, 1994

A;Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the

A;Reference number: A49963; MUID:94132002; PMID:8300571

A;Accession: A49963

A;Molecule type: mRNA

A;Residues: 1-639 <GUL>

A;Cross-references: UNIPROT:Q02817; UNIPROT:Q14884; GB:L21998

R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.

J. Biol. Chem. 267, 21375-21383, 1992

A;Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr

A;Reference number: A45106; MUID:93016075; PMID:1400449

A;Accession: A45106

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 626-1895 <GU2>

A;Cross-references: GB:M94131; NID:G186395; PIDN:AAA59163.1; PID:G186396

A;Note: sequence extracted from NCBI backbone (NCBIP:116706)

A;Accession: B45106

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 2037-3020 <GU3>

A;Cross-references: GB:M94132; NID:G186397; PIDN:AAA59164.1; PID:G186398

A;Experimental source: colon

A;Note: sequence extracted from NCBI backbone (NCBIP:116698)

R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M.

J. Clin. Invest. 88, 1005-1013, 1991

A;Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp

A;Reference number: A43932; MUID:91358717; PMID:1885763

A;Accession: A43932

A;Molecule type: DNA

A;Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>

A;Cross-references: GB:M74027; NID:G188863; PIDN:AAA59875.1; PID:G188864

A;Note: sequence inconsistent with the nucleotide translation

A;Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)

R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.

J. Biol. Chem. 264, 6480-6487, 1989

A;Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evide

A;Reference number: A33532; MUID:89197956; PMID:2703501

A;Accession: B33532

A;Molecule type: mRNA

A;Residues: 1916-2193 <GU4>

A;Cross-references: GB:M22405; NID:G188873; PIDN:AAA63334.1; PID:G188874

A;Experimental source: intestine

R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

J. Clin. Invest. 87, 77-82, 1991
 A:Title: Human bronchus and intestine express the same mucin gene.
 A:Reference number: A61257; PMID:91086481; PMID:1985113
 A:Accession: A61257
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 71, 1925-1948, 'TTS', 1952-1954 <JAN>
 A:Experimental source: bronchus
 R:Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Jones, C.; Forstner, Biochem. Biophys. Res. Commun. 183, 821-828, 1992
 A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-terminus
 A:Reference number: P00328; PMID:92198477; PMID:1550588
 A:Accession: P00328
 A:Molecule type: mRNA
 A:Residues: 2328-2468 <XUG>
 A:Cross-references: GB:M85523
 A:Experimental source: small intestine
 A:Accession: P00329
 A:Molecule type: protein
 A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
 C:Genetics:
 A:Gene: GDB:MUC2
 A:Cross-references: GDB:120203; OMIM:158370
 A:Map position: 11p15.5-11p15.5
 C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von Willebrand factor type C repeat homology
 C:Keywords: glycoprotein; intestine; tandem repeat
 F:2765-2834/Domain: von Willebrand factor type C repeat homology <VMC>

Query Match 22.0%; Score 1283.7; DB 2; Length 3020;
 Best Local Similarity 21.5%; Pred No. 6.3e-18;
 Matches 406; Conservative 97; Mismatches 498; Indels 883; Gaps 61;
 3 WK-----TLPIYLLLSVFIQVSSQELSCKGRCFESF-----E 38
 649 WREHVCKNDVSCPNQVFLYNLTTCQTCRSLSEADSHCLGEPAPVDGCGCPDHTLDE 708
 39 RGR-----ECQC-----DAOCKYKXC-CPDYESFCAEVH-----67
 709 KRCVPLAKSCYHRLYLEAGDVVVRQBERCVCRDLHCRQLRLIGSQCTAPKIMDC 768
 68 -NPTSPSSKAPPPSGASQTIKS-----TTKSPKPPN 100
 769 SNLTALATSK---PRALSCQTLAAGYHTECVSGCVCPDGLMDGRCGVVEKCP---821
 101 KKTKKVIESBEITEHSVSENQSSSSSSSSSTTWIKSKKNQAA-----149
 822 -----CVHNDLYSSGA-----KIKVDCNTCTCKRGRWVCTQA 854
 150 -----NREL-----153
 855 VCHGTCISYSGHYITFGKYIDFGHCSYVAVQDYCCQSSLSGSFSIITENVPCGTTGV 914
 154 -----QKLVKVNKK-----NRT 167
 915 TCSKAIRFMGRTELKDEKRVVIQRDEGHVAYTTREVQGLVWVESSTGIIIVMDKRT 974
 168 ---KKTTPKPPVDEAGSL-----DNGDFKV-----TTPD 156
 975 TVFKLAFSYKGV-----CGLGNFDRHSNNDFTTRDHVWVSSSELDFGNSWEKAPTCPD 1029
 197 TSTTQ-----HNKYSTSP-----209
 1030 VSTNPEPCSLNPHRRSWAEKQCSILKSVFISCHSKVDKPPFYACVHDSCTDGDCE 1089
 210 -----KITTAKPIN 218
 1090 CFCGAVASAOECTKEGACVFWRTFDLCPIFCDYNNPHECEWHYPCGNRSFETCRTIN 1149
 219 -----PPSLP-----224
 1150 GIHSNISVLEGVCYPCPKDRPIVEEDLKKCVTADKGCYVEDTHYPPGASVPTETCK 1209
 225 -----PNSDISKE-----TSLIVNKET-----241

Db 1210 SCVCTNSSQVVCPEBEGKILNQTDGAPCYWEICGNGVGVKHFNICSITTPSLTITFT 1269
 QY 242 -----TVETKETTNTKQSTIDGKEXTTSK-----267
 Db 1270 TITLPTPTPTSTTTTTTTTTSTVLSSTPKLCLMSWDINEDHPSSGSDGDRPFDDGV 1329
 QY 268 --EQSIEKTSKD-----279
 Db 1330 CGAPEIDERSVADPHLSLEQHQKQVQCDVSVGFICKNEDQFGNGPFGLCYDYKIRVNCC 1389
 QY 280 -----LAPTSKVLAKPTKAEITTKGPAITTKPEPTTPPKPEASTTKPKETPTTIKSA 333
 Db 1390 WPMDCIITTPSPPTTPSPPTTTTTLPTTTPSPPTTTTTTPPTTPSPPTTTTTPL 1449
 QY 334 P-TTPKEPAPTTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKAP-TTP 391
 Db 1450 PTTTPSPPTSTTTTPPTTPSP-PTTTTSPPTTPSPPT-TTTTTTPPTTPSPPTTP 1507
 QY 392 -----KEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEAP 443
 Db 1508 ITPPASTTTLPTTTPSPPTTTTTTPPTTPSPPTTTTTPTSTTTLPTTTPSPPT 1567
 QY 444 TAPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKAPTTKPEPAPTTKS 503
 Db 1568 TTTTPPTTPSPPTTTTTTPPTTPSPPTTTTTTPPTTPSPPTTTTTTPPTTPSP-PTTTT- 1624
 QY 504 APTTPKPSPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKAPPT 562
 Db 1625 -PTTP-----PTSTTLPTTTPSPPTTTTTTPPTTPSPPTTTTTTPPTTPSPPTTT 1679
 QY 563 APKEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTPEE 622
 Db 1680 TPSSPTTTTP--SPPT-----TWTTTPS--PTTTPSSPTTTTTTPSSSTTTP-SPPTTTMT 1730
 QY 623 PAPTTPKAAAPTKEPA-----PTTPKEPAPTTKPEPAPTT-----KETATTP--668
 Db 1731 PPTT-----TPSPPTTTMTLPTTTPSTLPTTTPSPPTTTTTTPPTTPSPPTTTTTTPPCV 1783
 QY 669 -----KG-----670
 Db 1784 PLCNWTGLSDCKPNFPGGDTLIGDVGCGPWAANI-SCRATMYDPVPIQGLQGVVCD 1843
 QY 671 -----TAPTILKEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTT 694
 Db 1844 VSVGLICKNEDQKQGVIPMAFLNYEVQCCVQPTTM---TTTTENPTPTPTT 1900
 QY 695 TT--TKEPTSTSDKP---APT-TPKGTAPTTPKEPAPTTTPKEPAPTTTPKGTAPTILKE 747
 Db 1901 TTTVTPTPTSTQSPNGLOAPTPTPTSTTTVTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 1959
 QY 748 PAPTTPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTT 801
 Db 1960 PPTPTGQPTTVLITTTTTMTPTPTPTSTKSTVTPTTTTTVTATPTPTGQPTPTM:PI 2019
 QY 802 --APTTPPTTPPTSEVSTPTTK-----EPTTIHKSPEDESTPELSABPTPKALNSPKE 854
 Db 2020 STTTVTPTPTTGTSTGTPHTSTAPAEULTSNPPESSTPQTSRSTSPSTESTILL 2079
 QY 855 PGVPTTKTAAKPEMTTAKDTERDLRTPTTTAAKMTKETAATTTKTTESKITA 914
 Db 2080 STLPAAIEMTSAPPSTPTAPTSTSGHTLSPPTSTTTPPTPTPTPTPTPTPTPTPTPTPTPT 2137
 QY 915 TTTQVSTTTQDTPKLT---LKTTLAKVTTTKTITITTEIMNKEE-----962
 Db 2138 STVQTTTSAWPTPTPLUSTSILHTGLRVPSSVLICCVLNDTYAPGGEVNGTVGD 2197
 QY 963 -----TAKPKDRATNSKATTPKPKTKAPKPTSTKPKTKMPRV 1002
 Db 2198 TCVFVNCSLSCLEFYNWSCPTSPPTTPSK-SFTPSKSPSTPSKPTPTGTPKPECPDF 2256
 QY 1003 RKPKTTPPKMTSTMPBLNPTSRTAEMLOTTTPRQTPNSKLVENP-----1051

Db 2257 DDPK-----QENETWLCDCFM-ATCKYNNVTEIVKVECEPPMPTCSNGLQ 2302
Qy 1052 ---KSDAGGA-----EGTPEHMLLRPHVFMFEVTPDM 1081
Db 2303 PVRVEDPDCCHWECDCYCTGWDPHYVTFDGLYSYCGNCTYLVLE-----EISPSV 2356
Qy 1082 D-----1082
Db 2357 DNFGVYIDNYHCDPNDKSCPRLLIVRHETQEVLIKTVHMPMVOVQVNRQAVALPYKK 2416
Qy 1083 -----YLRVFNQGIIN 1095
Db 2417 YGLEVYQSGINVVVDIPELGLVLS 2440

RESULT 2
T18262
S-layer protein - Clostridium thermocellum
C-species: Clostridium thermocellum
C-date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C-accession: T18262
R:Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
e.
A:Reference number: Z18847; MUID:93209931; PMID:8458832
A:Accession: T18262
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1664 <FUJ>
A:Cross-references: UNIPROT:Q06852; EMBL:X67506; NID:9236879; PID:9296881; PIDN:CAA47841

Query Match 19.6%; Score 1143.3; DB 2; Length 1664;
Best Local Similarity 21.2%; Pred. No. 1.6e-15;
Matches 367; Conserved 131; Mismatches 425; Indels 807; Gaps 66;

Qy 4 KTLPIYLLLLSVFV-----IQVSSQE 26
Db 6 KVLISILLTLTLLTSTSVNMSFAETPSIEMVLDKTEVHGVITATIKVNNIRLAGYQ 65
Qy 27 LSGKRCFSFGRGECDCQCKYDKCPDYSCAEVHNPTSPSPSK-----KAPPPS 82
Db 66 LNIK-----FDPEVLQVPDPAETGFTDKSP-- 92
Qy 83 GASQTKSTTKSPKPPNKKTKKVIETEEHSVSENQESSSSSSSSSSSTIWKIK 142
Db 93 -VNRVLLTNSKYGTP-----VAGNDIKSGIINFATGYNLLTAYK 131
Qy 143 SS-----PKPPVDEAGSLDN-----144
Db 132 SSGIDEHTGIIIGEIKVLKKQNTSIRFEDTSLMFGAISGTSLFDWDAETITGYEVIQPD 191
Qy 145 -----KNSAANRELQK-KLKYKD-----NKKQ-----RTKKKPT-----172
Db 192 LIWAEAPLKDASVALELDKTKVKVGDIIITATIKIENKMNKFNAGYQYQNIKYDPTMLEAIEL 251
Qy 173 -----PKPPVDEAGSLDN-----187
Db 252 ETGSAIAKATWPTVGTGLQSNYKGTAVANDVGAGIINFABEYNSLTKYRETGVAEET 311
Qy 188 -----GDFKVTPTD-----STQHNKV-----205
Db 312 GIIGKIGFRVLKAGSTAIRFEDTAMPGAIEGYMFDWYGENIKGYSVVQGEIIVAEGBE 371
Qy 206 -----STSPKITTAKINRPSLP-----PNS-----DTSK-----231
Db 372 PGEETEEVPTETPVDPPTVTEBFPVPSLPDSVIMELDKTKVKVGDIIITATIKIENM 431
Qy 232 -----ETSLVNNKETTETKETTITNKQTSIDGKEKTS-----265
Db 432 KNFAGYQNLNKKVPTMLEAIELETGSAIAKRTWPTVGTGTV-----LQSDNYKTTAVAND 486
Qy 266 -----AKETQSIKTSKOLAPTSKVLAKPTPKAETTT-----298

Db 487 VGAGIINFABEYNSLTKYRETGVAETGIIKIGKIFRVLKAGSTAI-----RFDTTAMPG 541
Qy 299 -----KGPALTTP-----KBPPT-TPKEPASTTPKEPTPT-- 328
Db 542 AIEGTVMFDWYGENIKGYSVVQGEIIVAEGBEETEBPPTETFDVDTPTVTEBFPVSELP 601
Qy 329 -----328
Db 602 DSYVIMELDKTKVKEGDIATIRVNNIKNLAGYQIGKYDPAKLEAFNIETGDPIDEGT 661
Qy 329 -----T 329
Db 662 WPAVGGTILKNRDYLPVTGVAINNVSIGILNFAAYVYVFDYRBEKSEDTGIIIGNIFRV 721
Qy 330 IKSAPTTPK-----EPATTTKSAFTTKBPAPT 359
Db 722 LKAEDTIRFEELESMPGSDGYMLDYNLRISGVVVIQPAIKAS-----DEIPTD 776
Qy 360 TKEPAPTTPKEPAP-----TTTKBPAPTTPKAPTTTPKAPAPTTPKAPAPT-- 413
Db 777 TPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 836
Qy 414 -----TPK-EPTPT-TPKEPAPT-TKEPAPTTPKEPAPTAPKKA-PTTPKEPAPT- 460
Db 837 PTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTSE 896
Qy 461 TPKEPAPTTPKPSPTTPKEPAPTTPKSAPTTTKAPATTTPKAPTTTPKPSPTTPKAPA 520
Db 897 TPPEPIPTDTPSDEPTPSDEPTPS-----DEPTPSDEPTPSDEPTPSDEPTPSDE 953
Qy 521 PTPKEPAPTTPKKA-----PTTPKEPAPT-TPKEPAPTTPKAPAPTAPK 565
Db 954 PTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTSD 1013
Qy 566 EPAPTTPKEPAPTTPKKTTPTEKLAPTTPEKAPT-----TPRELAPTTPEEPTPT-TPE 621
Db 1014 EPTPSD-----EPTPSDEPTPSDEPTPSD-EPTPSDEPTPSDEPTSE 1052
Qy 622 EPAPTTPKAAAPNTKEPAPTTPKKA-----EPTTPKEPAPT-TPKETAPTTPKAPTTLKE 678
Db 1053 EPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 1112
Qy 679 PAPTTPKAPKAPKELAPTTKEPT-SKTSKAPATTTPKGTAPTTTPKAPAPTTPKAPAPTTP 737
Db 1113 PTPS--DEPTPSD-EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS--DEPTPS-- 1165
Qy 738 KGTAPTTLKEPAPT-TPKAPKAPKELAPTTKGTSTTSDKAPT-----TPKETAPTTPKEP 793
Db 1166 --DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD-EPTPSDEP 1217
Qy 794 APT-TPKAPAPTTPPTTPPTSEVSTPTTKETTHKSDEPTSELSAETPKALENSP 852
Db 1218 TPSETPEEPTPTDTPSDEPTPSD-EPTPSDEPTPSDEPTPSDEPTPSDEPTPS--SETP 1266
Qy 853 KEQVPTTKTAPKAPKEMTTAKDKTERLRTTPETTTAAPKMKETATTTKETSKEI 912
Db 1267 EEP-IPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1325
Qy 913 TATTTQVTSITQDTPFKITTLTKTTLAPKVTTKTITTTIMKNKEETAKPKRATN 972
Db 1326 PTDTPSDEPTPSDEPTPSDEPT-----PSDEPTPSDEPT 1360
Qy 973 SKATTPKQ-KPIKAPKPTSTKKPMTMPVRKPTTTPPKMTS-----1016
Db 1361 SDEPTPSDEPTPSDEPT-----PTTPTPTPTPTPTSGSGSGSGGG 1407
Qy 1017 -----TWPELNPTSRIAEAMLQTTTRNQTNPNSKLVNPKSEDAAGAEPTPHMLL 1068
Db 1408 GGGGGTVPSTPTPTPTS-----KPTSTPAPEIE-EPTPSDVPAIGGEHAYL 1455
Qy 1069 RPH--VFMEP-----1076

Db 1456 RGYPDGSPERNITRAEAAVIFAKLLGADSEYGAQASPYSLDLADTHAAWAIAKEATSQ 1515
 QY 1077 -----VTPD-----MDYLPRVNOGI-----IINP 1096
 Db 1516 GLFKGYPDGTPFQDNITRAEATVVLHFLTKVKGQEIMSKLATIDISNP 1565
 RESULT 3
 T16251
 hypotrichal protein F35A5.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
 C:Accession: T16251
 R:Leimbach, D.
 submitted to the EMBL Data Library, January 1996
 A:Description: The sequence of C. elegans cosmid F35A5.
 A:Reference number: Z18485
 A:Accession: T16251
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1274 <LEI>
 A:Cross-references: UNIPROT:Q02007; EMBL:U46675; NID:g1166613; PID:g1166621; PIDN:AAB526
 A:Experimental source: strain Bristol N2; clone F35A5
 C:Genetics:
 A:Gene: CESP:F35A5.1
 A:Map position: X
 A:Introns: 1272/2
 Query Match 17.4%; Score 1015.5; DB 2; Length 1274;
 Best Local Similarity 23.8%; Pred. No. 3.6e-13; Mismatches 364; Indels 629; Gaps 76;
 Matches 343; Conservative 107;
 QY 77 KAPPSGASQITKSTKSPKPPNKKTKVIESEITEEHSVSENQBSSESSSSSSSS 136
 Db 3 RAPPTP-----IKNPAKW-KPP-----WESVDEEEMEVEDETPAPS----- 39
 QY 137 TIWKIKSSKNSSANRELOKK-LKYNKNKRTKKTKPKP----- 175
 Db 40 -----KLEKSLKRRK-----APTKEVPSGAPSPVPIKPVKWKWAPW 79
 QY 176 -----PVVDEAGSLDNGDFKVTTPDTSTTGHNVSTSPKITTAKPINRPSL 223
 Db 80 EDDPMEEAAPVPAK-----KVRDP-----SPKKVPKPRDASP-- 115
 QY 224 PNSTSKESTLVNKKETTETKTTNKOTSTDGKTKTSK-----ETQSIKTSADL 280
 Db 116 -----KKIVAAKKEPETLPAVPPTPVKNPVKKFKAPWEDDEVVDVKD- 159
 QY 281 APT-----SKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEP-ASTTP----- 322
 Db 160 APTVPAKTPVLKKEPEAAAKPRDP-----SPKKAAPSKHEHPVIPPPIKPKAKWKPPW 216
 QY 323 -----KEPTPTTIKSAPTTKPEP-----PTTIKSAPTTKPEP--APTITKE 362
 Db 217 EDDEVPTBEIKEPEPATRKVPALKKKKPSSTSVKPSVSPSPTKVVP-VKKEVPEVPTPIKN 275
 QY 363 PA-----PTTKPEAPT-TTKPEAPT-TTKPEAPT-TTKPEAPT-TTKPKPA 403
 Db 276 PTKWKWKPPWEDETVEEVEKPEVPEKAPVLUKKDDPAPAAKARDPS-----PSKAAAPKKVE 331
 QY 404 PTTKPEAPTTPKEPT-----PTTKPEAPTTPK-----EPAPTTP-K 439
 Db 332 PSSPVPV-PTPVKNPVKKYKPPWEVDDBPAAEVKKPSAPEKKTVPVLRKKEPESSTTPSS 390
 QY 440 EPAP-----TAPKKAP----- 451
 Db 391 DFSPKKAAPVAKPRDSSPKKATPLQADPKAGEVPTPVKNPVKKYKPPWEVDDSDPVEEV 450
 QY 452 TTPKPEAPTTP-----KEPA-----PTTTKEP-----S 474
 Db 451 KOPEAPAKKTVPVLRKEKEFAAKDTAKFATSKTPTPEKDPVKKPRDSSPKKVAAPKPSAQ 510
 QY 475 PTPP-KEPA-----PTTTKSAPT-TTKPEAP-----TTTKSAP 505

Db 511 PATPVKNPVKKWRPPWEDDETADOVSKETDAKTFSLAKKDPAPAKESLKPKAUTK-AP 569
 QY 506 TTPKESPTTTKEPAPTTPKPEAPTTPPK-----PAPTTPKPEAPTTPKPEP----- 551
 Db 570 AKBRDPS---KKVAPTAEKKTVPVLAKEPEAPADSKTKEPEKSPRDPSPKAVPAKP 626
 QY 552 -----APTITTKKAP-----TAPKEPAPTTPKETATPTP-----KKLTP----- 585
 Db 627 VPKEVAPAAVKAPEDISKPKDTAPKAEPSG---VVPPTPVKNPVKKWKPPWEDDDAPA 684
 QY 586 -----TTPSKLAPTTPKEKAPTTP-----ELAPTTPSEPTP--TT 619
 Db 685 KPVSLPEPEKKTVPVLAKE-APTKPDSEAAADPVSGPSSKDKPLAKKAPVKPRDPSPMKAV 743
 QY 620 PEEPAPTT-----PK---AAPNTKPEAPTTPKPEP----- 647
 Db 744 PIKPAKTEVPPAVVKKPEPVAKSRDPSPKKAKAEPSGVP-PTPVKNPVKKWKPPWED 802
 QY 648 -----APTTPKPEAPTTPKETATPTTPKGTATPTTPKGTATPTTLKEPAPT- 683
 Db 803 DDAPAEVAVNPBEKKTVPVLAKEKTPVKPRDPSKKAAPKAPSTKTDAPVSVKKPEPVSK 862
 QY 684 PKKAPKE-----LAPTTTKEPT-----STSDKPAPTTPKGTAPT-TPKEP 724
 Db 863 PKPEPSKKAEPNSPVVPTPVKNPVKKWKPPWEDDEPTVEEVKKEPSEKKTVPVLAKEP 922
 QY 725 -----APTTPKPEAPTTPKGTATPTTLKEPAPT-----TPKKPAP-----KELAPTTPK 767
 Db 923 EXPKDAKPAKAPRDPSPKKAYPE--KEPAKVAAPKPRDLSPKKAIPIANTQEAAPTVPK 980
 QY 768 GPTS-----TTSKDP-----APTTPKPEP-----TTPKKPAPTTP 806
 Db 981 NPVKWKPPWEDDDEPAEPVSAPEPEKKTVPVLAKEKAPKPRDPSKKAAPVAAKDPDKIP 1040
 QY 807 ETPPTTSEVSPNTTKEPTTIHKSP-----DSTPELSA-EP---TPKALENSKPEGV- 857
 Db 1041 EVP-----PTPVKNPVKKWKPPWEDDDDEPSPVSAPEPEKKTVPVLAKEKAPKPKATK 1091
 QY 858 -----PTTKTAPAT-----KPEMTTAKDKTTERDLRTTETT--TAAPKWT 897
 Db 1092 PDSEAAADPVSGTSPKSLSKKAPVEKPKPTDPKDKLKPSPAKKPEKAPAPAAKWK 1151
 QY 898 KETATTTKTESKITATTQVSTTQDTTTFKLTLLKTTTLAPKVTITTKTITTEIM 957
 Db 1152 KPVWDDDPDEADFTVPAPSKKPDTEADPLG-----GPKTKDPK-----L 1194
 QY 958 NKPEETAKPKDRATNSKATTPKPKPTKAPKAPTSTKPKTMPRVRKP----- 1006
 Db 1195 NKKAPEAKPTEK-----PKPEVSKPEPPKPTPEPKP-AAPKKWKPPWEDDDPEAD 1245
 QY 1007 -TTPTRKMTSTWPELNPSTRIAEANLQTTTRPNQTPNSKLVENPKSED-----AGGAE 1061
 Db 1246 FTMAPKK----- 1268
 QY 1062 ETP 1064
 Db 1269 KDP 1271

RESULT 4

T30826
 nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
 N:Alternate names: alpha-NAC protein
 C:Species: Mus musculus (house mouse)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T30826
 R:Yctov, W.V.; St-Arnaud, R.
 Genes Dev. 10, 1763-1772, 1996
 A:Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle-
 A:Reference number: Z20889; MUID:96312450; PMID:8698236
 A:Accession: T30826
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-2187 <YOT>
A:Cross-references: UNIPROT: P70670; EMBL: U48363; NID: G1666688; PID: G1666688; PIDN: AAB187
C:Gene: Naca
A:Map position: 10
A:Introns: 24/1; 1996/1; 2050/3; 2050/3; 2099/3; 2142/3; 2183/3
A:Note: differential splicing converts alphanac into a tissue-specific DNA-binding active
C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match 17.4%; Score 1014.9; DB 2; Length 2187;
Best Local Similarity 19.3%; Pred. No. 7.9e-13;
Matches 375; Conservative 147; Mismatches 426; Indels 991; Gaps 73;

QY	21	QVSSQ-ELSCGRCFESPERGECDCDAQCKKYDKCCPD-YESFCAEVHNP-----	69
DB	183	QVFSQGTNLKGA-----PCPDVVRAPFSLHLENPLASVQPGLM	222
QY	70	-----TSPSSKKAPP-----	80
DB	223	SCPQTLNNTSPKGVPISSALTQSRSLNLKGFVSPPARNTAAPSILAPSTSLGCHLPL	282
QY	81	-----	80
DB	283	LHSSVDSPIQPGQSGILAVNPTSVGHSGTAASCPCPCVVPALPRLLAVDGAAPSD	342
QY	81	-----PGASQTIKSTTKRSPK-----PNKKTKKVI	108
DB	343	DKGSAVTNELCSPPGSSNV--AGTSLSPKASLVKGSNVALQLVLTQVPASQKTGLKEI	400
QY	109	ESEETEEHVSSEN-----	122
DB	401	PVSCIGATHALDNPSAISVAPATHVPPPTSSGLVSKDPAFVTSVLVVAHKKQFPAPP	460
QY	123	-----QSSS	126
DB	461	ASATLGVFVPLPATEGLKNLPI SALVNVGAPVSPAQAGLPTRKDTLLQPLAPIALKESP	520
QY	127	SSSSSS--SSSTIWK-----	140
DB	521	SSQSASSLEVLSEDIVTKTTGGPAPVVRPAIAGVATTSLRADSPFAVIRADSCVSPNT	580
QY	141	-----IKSKNSAANR-----	151
DB	581	VSQPLKRSVTDPAAPRAKNTAPSTTSLVPLASEGCPVASSMALSQNASVSETALAL	640
QY	152	--ELQKLKVDNKNRKKKFTPKPPV-----VDEAGSLDNGDFKVTTPDTST	199
DB	641	SPEIPKSV-----PFPDPLAEISFSNARKVDVSHMESGSSGRQGHDPASV	687
QY	200	TQH-----NKYST	207
DB	688	TAKGTVVCLADSSLDTSVSASGSLASGASSPLYPLEVSFUEAGLAVQGPKGSLNKLSP	747
QY	208	SPKITTAKPINRPSLPPNSDTSKETSLSLVNKKETTVEK-----	247
DB	748	TPPSSKAPV-FSTGAPP---SPKGAPIVPTSEISSKQVPAEILPSPQKTPVETASRL	802
QY	248	-----TTTNKQSTDSGEKKTSAKTCSEKTSQ-----KDLAPT	283
DB	803	ISAVQSPKVDPIMSDVPTSPKXTSATAVPKDTSA--TSLSKSVPAVTSLSPPKAPVAPS	860
QY	284	SKVLAKP-----TPKAEITTKGPALT-----	305
DB	861	NEATIVPTIELTSLNALAAATPKETLATSIPKVTSPSPQKTPKSVSLKGAPAMTSKAT	920
QY	306	-----PKE-----PPTTPKFPASTT-----PKEPTTTTKSAPTT	336
DB	921	ELAASKDVSPSPKPEVPILOHVPTSPKSPVSDTLGALTSPPPKGP-PATLAETPT	979
QY	337	PKE-PAPTTTKSAPTTK-----EPAPTTTKEPAPTTKBPAPTTTKEPAP-----	381
DB	980	PKKSPKPAASKKTPATPSPGVTAVPLEIPCKCKKAPKTAAPKESATSSSSSRAPKPTAVS	1039

QY	382	-----TTTKSAPTTKBPAPTTPKK-PAPTTPK-----	409
DB	1040	KEIPSKGVTAIVLEISLPLKETSASA--TPGEKASSPKRSKPTAGKPTPPGGVTAIVPP	1097
QY	410	-----PAPTTTKPKEPT-----TTPK-----PAPTTKE	432
DB	1098	EISLPPKRETONATPNESLAASQKRSKPTSVKETPPGGVTAMPLEIPSAQKAPKTAV	1157
QY	433	P-----APTTKBPAPTAAPKAPTT-----	454
DB	1158	PKOIPTEDAVITLAGSPUSPKKASTAAPKEAPATPSVGVIAVSEISPSPKKTSKATA	1217
QY	455	-----KBPAPTTKBPAPTTKE-----PSPTTPK-----PAPTTKSA	490
DB	1218	PKENSATLPPKRSKTAAPKE--TPAISSEGVTAIVSEISPSPTPASKGVVLTLPKGA	1276
QY	491	TTTKBPAPTTTKSAPIT--PKES--PTTKBPAPTTKBPAPTTTKKAPTT-----	539
DB	1277	NALAE-SPASPKVKPTAAPEETSTTSPQKIEKVAGPKESATSPSKKTPKTAIVPKETS	1335
QY	540	-PKBPAPTTKBPAPTTTKKAPTAAPKE-PAPTTKBPAPTTTKKLTPTTPEKLAPTTPE	597
DB	1336	APSEGVTAIVLEIPSPKAPKTAAPKETAPS--PEGATTAIVQI-PPSPRKGSKKAGS	1392
QY	598	KPAPTTPEELAPTTPEEPTTPEEP--APTTPKAAAP-----NTP	636
DB	1393	KETPTTP-----SPEGVTAAPLEIPISSKKTSMASPKETLVTPSSKLSQTVGPKETS	1446
QY	637	KEBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP	694
DB	1447	LEGATAVLEIPSHKAPKTVDPKQVLTTPSPKADPTTLAE--SPSPK-APKTAAPPS	1504
QY	695	-TTTKPTSTTSKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP	743
DB	1505	ERVTTVEP-----EKPA-TPQKASGTTASKVPVPAETQEVAVSSRETPTVPAVPK	1558
QY	744	TLKBPAPTTKBPAPKELAPTTKGTSTSDPAPTTKBPAPTTKBPAPTTKBPAPTTKBP	803
DB	1559	SHKTSKTIKBPAPTLPSPTKSPKIPSSKK-----APTSAP-----KEFP	1602
QY	804	TTPTPTPTTSEVSTPTTKPTTIHKSDESTPE--LSAETPKALENSPKPGVPTTK	861
DB	1603	ASPSIKVTTSLAQ-----APSLQKAPSTTIPKENLAA--PAVLVSSKSPAAP-AR	1653
QY	862	TPAATKPEMTTAKDKTTERDLRTP-----ETTTA-----APKMTKETATTT	904
DB	1654	ASASLSP---ATAAPQAPKEATTIPCKKAAATETPIETSTAPSLEGAPKETSETSV--	1708
QY	905	EKITESKITATTQVTSITTTQDTPPKITL---KTTTLAPKVTITTKTITTEIMNKP	960
DB	1709	-----SKVLMSSPPKASSSKRASTLPATTLPSLKEASVLSPTATSSGK-----	1752
QY	961	EETAKPKDRATNSKATPKPKPKPKPK-----	989
DB	1753	DSHISPVSDACSTGTTT--PQASEKLPSKKGPTAFTMLAAPAPEASALAITAIQKSPGA	1810
QY	990	-----PTSKKPKTMPR--VRKPKTTTTPKMTSTMPELNPTSEIAEMLOTT	1035
DB	1811	NNSASSPKCPDPSKSKDTKGLPSAVALAPQTVPEK-----DTSKALETLLVSP	1860
QY	1036	TRPNQ-----TPNSKLV	1047
DB	1861	AKGSDCLHSPKPGVSGQVATFLAAFTSDKVPPEAVSAPKAPAPASALTLAPSPVAPLP	1920
QY	1048	EVNPKSDAGAGETPHMLLRPHVFMPEVTPDMDVLPV-----PNOGIIN--	1095
DB	1921	PKOPLLESAPGSLVLESPSKL-----PVPAEDELPLIPPEAVSGGEPFQFILNMP	1972
QY	1096	-----PML 1098	
DB	1973	APKPACTPAPAPSAKQPVL 1991	

RESULT 5

T31108
 cyst germination specific acidic repeat protein precursor - Phytophthora infestans
 C:Species: Phytophthora infestans (potato late blight agent)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T31108
 R:Goernhardt, B.
 submitted to the EMBL Data Library, April 1998
 A:Reference number: Z20986
 A:Accession: T31108
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1489 <GB>
 A:Cross-references: UNIPROT:O96449; EMBL:AF061185; NID:G3851513; PID:G3851514; PIDN:AACT
 C:Genetics:
 A:Gene: car90

	Query Match	17.3%	Score 1009.9	DB 2	Length 1489
	Best Local Similarity	27.0%	Pred No. 5.8e-13		
	Matches 406	Conservative 65	Mismatches 474	Indels 561	Gaps 69
QY	27 LSCGRCFESFERGEC-----DCDAQCKYD--KCCPDYESFCAEVN-----	68			
DB	43 IPCSGVGAEPV--GTACPKAGDVATSDCQPYLLSYNGAVCVAPVDACALIHDDMWCCEP	100			
QY	69 -PTSPSSKKA-----	78			
DB	101 PKGTSAVEAEETIAAYNGESSGWTGHDEVVQGVDBEEIPARVNDTVDDPIGNCE	160			
QY	79 -----PPPSG-----ASQIKSTTKRSPKPPNKKTKKV	107			
DB	161 VATEATQGHATEGGKYDYDTSTGTQCDYGNTHYGSTTTEGVTGGYGTGTD---AKV	216			
QY	108 IESE-----	111			
DB	217 IDGETYLDYPTGITEIIEEDGTPPGYGTDDGGTTGGYTVTDNTHETGAGGYDAGTR	276			
QY	112 -----EITEHVSSENOESSSSSSSSSSST-----IWKIKSSKNSAANRE	152			
DB	277 BEYSTPTVGVSTETEGCHVTGGVPSDETEATEGTGYVPRSETTAAPSEDITYAPRE	336			
QY	153 LQKLVKDKNKRKKPTPKPPVVDGAGSLNG-----	189			
DB	337 V-----TPYAPTEKPYDVEETTYVTSESTYAPTKSETNAPTERMHYAHIEKP	383			
QY	189 -DFKVT---TPDSTTQHNKVSTSP-KITTAKPINRPSPPPNSDTSKTSLVNKEIT	242			
DB	384 CDEVTWYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE	438			
QY	243 VETKETT---TNKQTSIDGKETSAXETOSIEKTSAKDIAPTSKVLAKPTPKAETTK	299			
DB	439 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY	495			
QY	300 GPALTTPKEPTTPKEASTPKPEPTTTPKSAPTTPKEAPT-TTKSAPTTPKEAPT	358			
DB	496 ASTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY	554			
QY	359 --TTKEAPTTPKEAPTTPKAPTTPKAPT-----TPKEAPTTPKAPTTPKAPT	406			
DB	555 BETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY	613			
QY	407 PKEAPTTPKEPTTPKPE-----AP-----TTKEAPTTPKPE-----AP	443			
DB	614 PYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE	673			
QY	444 TAPKAPATTPKEAPTTPKAPTTPKAPT-----APT-TTKSAPT---	491			
DB	674 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY	733			
QY	492 -----TTKEP-----APT-TTKSAPT--TPKEPSPTTKPE-----APTTPKEP	529			
DB	734 MYAPIETTYGTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYGTEETTY	793			

RESULT 6

I38346
 elastic titin - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004
 C:Accession: I38346
 R:Leibit, S.; Kolmerer, B.
 Science 270, 293-296, 1995
 A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
 A:Reference number: A57430; MUID:96026330; PMID:7569978
 A:Accession: I38346
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-7962 <RES>
 A:Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g101
 C:Genetics:
 A:Gene: GDB:TTN
 A:Cross-references: GDB:127867; OMIM:188840
 A:Map position: 2q31-2q31

QY	530 TTPKAPAPTPKAPAPT--TPKAPAPTTPKAPAPTAPKAP-----APTTPK	573
DB	794 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY	853
QY	574 ETAPPTPKLTTTPEKLAPT--TPKAP-----APTTPKELAPT-----	610
DB	854 TYAPTEKTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE	913
QY	611 -----TPESPTPTTPEP-----APTTPKAAAPNTPKAPAPT--TPKAPAPT	651
DB	914 TEETTYAPAEETPYSPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY	973
QY	652 PKEP-----APTTPKAPTTPKAPTAP-----TTLKEPAPTTPKAP-----	689
DB	974 PYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP	1033
QY	690 KE---LAPT-----TTKEPTSTTSCKPA-----PTTPKGTAPTTPKAPAPTTPK	730
DB	1034 TEETTYAPTEETTYASTBETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTY	1093
QY	731 EPAPITPKGTAPT--TLKEPAPTTPKAP-----APKE---LAPT-----	769
DB	1094 TYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAP	1153
QY	770 TSTTSCKPAPTTPKGTAPT--TPKAPAPTTPKAP-----APTTPETPTTSEVSTPT-	820
DB	1154 TEET--YAPTEATTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE	1211
QY	821 -TTKEPTTHHKSPDSTPELSAETPKALENSPKPGVPTTKTAAKDEMTHAKDKYT	879
DB	1212 ETTIYAPT--EETPYEPTTEETTYAPT-----EETPYEPTTEETTYAP-----	1250
QY	880 ERLDRITPTTTPKAPMTKET-----ATT--TEKTTESKITATTQVTSST-----	926
DB	1251 EETTYAPTEETTYAP--TEETTYAPIDETTYGTEETTYAPTEETTYAPTEETTY	1308
QY	927 TT--PFKIT--LKTTLAPKVTIT-----KKTITTTTEIMNKPEETAKP	966
DB	1309 TTVEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP	1368
QY	967 KDAATNSKATTPKPO---KPTKAPKKPTSTKPKTMPVRKP-----	1005
DB	1369 TDEPTDEPDEPTDEPTDEPTDLPT-----TDEPST-----PCDNQINGIGVENKVRYN	1417
QY	1006 ----KTTPTPRK-----MTSTWPELNPTSGRIABAMLQTTTR	1037
DB	1418 NAGIYNTTPGPRNSQSWHSCCRSCYNDPICHAFSPHQTS-----SDSVCELTTSTSDR	1470
QY	1038 PNQTPN 1043	
DB	1471 BEDQQN 1476	

Query Match 17.3%; Score 1008.4; DB 2; Length 7962;
Best Local Similarity 15.4%; Pred. No. 6.6e-12;
Matches 385; Conservative 126; Mismatches 460; Indels 1536; Gaps 90;

QY 2 AMK-----TLPIVILLLLSVFVIOVSSQ-----25
Db 5411 AWEHLQDVTLEGGTCMTVO-----FSVNVKSEWFRNGRILKPGGRHKTEVEHKV 5463

QY 26 -----ELSCGRGCFSEFGRGECDAQCKYDK-----CCPDYBSFCAE- 65
Db 5464 HKLTADVRAEDQGYTKC---YEDLETSABLRIEABPIQFTTKRIQNIWSEHQSATFEC 5520

QY 66 -----VHN-----68
Db 5521 EVSFDALVWYKGTPELTESQKYNFRNDGRCHYNTIHNTPDDDEGVSYVARLEPRGEA 5580

QY 69 -----PTSP-----PSK 76
Db 5581 RSTABLYLTKEIKLELKPDPIDSRVPIPTPIRAVPESIPVAVPVLPLPTPEK 5640

QY 77 KAPP-----PSGASOTI-----88
Db 5641 KPPPKRIEVTKAVKDAKVAKPKEMTPREBIVKPKPPPTTLIPAKAPBIIDVSSKAE 5700

QY 89 -----88
Db 5701 EVKIMTIIRKKEVQKEKAVYKKAQVHKVKEKVFIESPEBPYDELEVEPYTEPPEQPYE 5760

QY 89 -----88
Db 5761 EPDEYEEIKVEAKVHEWEEDEPBGQYVYEREGYDEGESEWEEAYQEREVIOVQKE 5820

QY 89 -----KSTTKSPKPN--KK-----KTKKVISEBI-----113
Db 5821 VYESHKRVPAKPEKKAPPKVIKPVIEKIEKTSRRMEEEKVQVTKVPEVSKKIVP 5880

QY 114 -----TEHSVSENQSSSSSSSSSSSTIWKI-----141
Db 5881 QKPSRTVQEEVIEVKVPAVHTKQVISEKMFASHTEEBSVTVPEVQKEIVTEEKIH 5940

QY 142 -----KSNXNSANRELQKKLVKONKNTTKKP-----171
Db 5941 VAVSKRVSEPPKVPPELPEKPAPEAVPVPVPPKKEPPAPKVPPEVPPKPPVPEKFPVPVK 6000

QY 172 -----TPKPPVDEAGSLDNGDFKVTTPDTSTTO-----HNKYSTSPKI 211
Db 6001 KEPAAPPKVPVPPKVPPEE-----KLPVPVAKKEAPPKAPKEVQKGVVTEKI 6050

QY 212 TTA-----KPINR-----PSLP-----224
Db 6051 TIVTQRESEPPPAVPEIPKKKVPPEBKVP-PRKEEVEPPPPKVPALPKKVPPEEKVAVPV 6109

QY 225 -----PNSDTSKET-----233
Db 6110 PVAKKAPPBRAVSKKVVVEKRFVAEKLSTFAVQORVEVTHVESABEESYSEBEGV 6169

QY 234 SUTV-NKETVETKETTNTKSTDGKETTSAK-----ETQSIKET-----275
Db 6170 SISVYREEREEREAEVTEYVMBEPEBYVVEEKHLIISKEVAEPAEVTEVQEKIYL 6229

QY 276 -----SAKOLATSKVLAKPTKAEITTKGPALTTPKPTPTTPKEP 317
Db 6230 KPKIPAKIEBPPPAKVPPEAKKIVPEKKVPA-VVPKKE-----KVPPPKVPEEP 6277

QY 318 ASSTP-----KEPTP-----TTIKSAPTPKPEPAPTTKS 347
Db 6278 KKPVEKKVPPKVIKWEELPAKVTEKMQIQTQEKVLVAVTKKEAP--PKARVPEEPKR 6335

QY 348 A-----PTPKPEPAPTT-----KEPAP-----365
Db 6336 AVPEEKVLKLPKREBEPKAVTEPRKRVVKEBKVSIEAPKEPQPIKEVTIMEKERAY 6395

QY 366 -----TPPK 369
Db 6396 TLSEEAHSVQREBEYBEYDYKEFEFEYPTBEYDQYEEYERERYEHEEYITEPE 6455

QY 370 EPAPT--TTKEPAPTTTKSAP-----TPPKE--PAPTTPKPAPTTTPKEP-----410
Db 6456 KPIPVKEVPEPPTPKAPPAKVLKAVPEEKVPVPIPKLAPPPKVPPEEKVPEEK 6515

QY 411 -----APTTPKE-----417
Db 6516 IHISITKREKEQVTEPAAKVPMKPKRVVABEKVPVPRKEVAPVPRVPEVSELEPEVAP 6575

QY 418 -----PTPTTPKBPAPTTTPKEP-----439
Db 6576 EEEVTHVBEYLVEESEEYIHSEEEITBEEVVPVVPKVPPEPRKVPPEKAPVPPVK 6635

QY 440 EPAPTA-----PKKP-----APTTPKEPAPTTTPKEPAPTTTKPS--474
Db 6636 KEAPPAKVPPEKPKPEKVPVLIPKKEKPPPAKVPPEKPPV-EEKVPVPEKVPPEAPP 6694

QY 475 -----PTTPKEPAPTTTKSAPTTKEPAPTTTKSAPTTTPKE-----PSPTTKPAPT 522
Db 6695 AKVPEVKKEVPPEKVPVPAKKEVAPPA--KVPEVPPKLIPEEKKPTVPVKKVPEAPPK 6752

QY 523 TPK--EPAP-----529
Db 6753 VPKGEVVPVVALPQEEEVLPQEEIVPEBEVLPPEEEVLPPEEEVLPPEEEI 6812

QY 530 -----TTPKKPAPT-----538
Db 6813 PPBESEVPPBESEVPPBESEVPPBESEVLPPEVKVPVPAVPVPEIKKKVTEKKVVPKKEA 6872

QY 539 -----TPKE--PAPTT-----PKEP-----551
Db 6873 PPAKVPPEKKEKRIILPKBESEVLPVEVTESEPEEPISESEIPEEPPSEIEVEEVAPP 6932

QY 552 -APTTPKAPAPTAP-----KEPAPTTTPKET-----575
Db 6933 RVBEVIKAVPEAPTVPVKVVEAPPKAVSKKIPEEKVPVPOKKEAPPKAVPEVPKVPKE 6992

QY 576 -----575
Db 6993 KXVLVPKKEAVPPAKGTVLLEKVSVAFRQVNVVGRLELVVEAEVEBIEPEEEFHEVE 7052

QY 576 -----575
Db 7053 EYFEEGEFHEVEFIKLEQHRVBEHEHVEKRVHVEVFEAESEVEVEKPKAPKGPSE 7112

QY 576 -----APTTPKCLT-----PT-----TPE 589
Db 7113 KIIPKPPKPTKVPPEKPPPAKVPPEVKKIYVBEKRVPEPEPRVPTKVPPEVLPPEKVPPE 7172

QY 590 KLAPT--TPKEPAPTTPEELAPTTPEE--PTPTTPKEP--APTTPKAAAPN--634
Db 7173 KKVPPPAKPEAPPPKVPPEAPKEVPEKVPVPPKPPKPPVPTKVPPEVKAAPVEKKVP 7232

QY 635 -----TPKEPAPTTTPKEP-----APTTPKEP-----APTTPKETAFTTPKGTAPTTLKE 678
Db 7233 EAIIPKPESEPPPEVEPEEPPSAP--PKKPEVPPVPRVPEVPEVPEKVPRAAP-----7285

QY 679 PAPTTTPKP-----APKELAPTTKPTSTSDKPAFTTPKGTAPTTTPKEP-----724
Db 7286 -----PKKPEVPPVPEAPKEVVP-----EKKVPVPP-----PKKPEVPPFT 7322

QY 725 -APTTPK-----EPAPTTTPKGTAPTTLKEP-----AP-----T 751
Db 7323 KVEVPKVAPEKVPPEAIIPKPESEPPPEVEPESEVPEEPAEVEVEPEEPAAPQVTV 7382

QY 752 TPKKPAPKELAPTTTKGPTSTTSDKP-----APTTPKETAP-----TTPKEP-----793
Db 7383 PPKNPVPEKKAP-----AVVAKKPELPPKVPPEVPEKVPVPEKVPVPEKPEAPPAK 7435

QY 794 APTTPK-----PAPTTTPETTPPTTSEVSTPTTKE-----PTTIHKSPPD 833

```
Db      7436 VPEVPEKVVPEKVAVPKPPVPAKVPVPEKVPVLEKPAVPVPEKVAESPPPEVVEPE 7495
Qy      834 ESTPE-----LSAETPKALENSKEPG-----VPTTKPAATKPEMTTAK--- 875
Db      7496 EIAPEEETAPBEKEVPVVAEEBEPVPPPAVPEPKKIIPBEKKVPVVIKKPAAPPKPEP 7555
Qy      876 DKITER-DLRTTPTTTAAP-----KMTKETATTTTEKTTESKITATTT 917
Db      7556 EKVIEKPKLKRPPPPPPAPPKEDVKEKIFOLKAIPKPKVPENQVPEKV----- 7605
Qy      918 QVTSITTQDTTFFKIT--TLKTTTLAP-KVTTTKTITTTTMMKPBETAKPK-----D 968
Db      7606 -----ELTFLKVPGGKEKVKRLLPKPKPEKPEEVVLKSVLRKPEEBEPKPEPKLE 7657
Qy      969 RATNSKATTPKQPKTKAPKPTSTKKPKTTPRVKPKTTPTRKMTSTMPELNPTGRI- 1027
Db      7658 KVKKEAVPEPPPKPEVEEVEVPTVTKREKIPETK-----VPEIKPAIPLP 7704
Qy      1028 -----AEAMLQT-----TTRPNQTPNSKLVEV-----NPKSEDA----- 1056
Db      7705 APEPKPEAEAKTIKPPPEPEPTPIAAVTPVVGKKAEEKAKPEAAKPKGPIKGV 7764
Qy      1057 -----GGAEGTTPHMLLRPHVFMPEVTPDMYDLPRVP 1088
Db      7765 KKTSPPIAEARRKLRPGSGGKP-----PDEAPFTYQLKAVP 7801

RESULT 7
T45025
mucin MUC5B, tracheobronchial (imported) - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T45025
R:Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A:Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
A:Reference number: Z22899; MUID:97166151; PMID:9013550
A:Accession: T45025
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3570 <DES>
A:Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
A:Experimental source: placenta
C:Genetics:
A:Gene: MUC5B

Query Match      17.0%; Score 994.8; DB 2; Length 3570;
Best Local Similarity 13.5%; Pred. No. 4e-12;
Matches 445; Conservative 115; Mismatches 459; Indels 2282; Gaps 98;

Qy      6 LPYLLLLLVFVIOQSSQELSK-----GRCFESPER----- 39
Db      3 LPV-----STVCREV-----CAWSWYNGHRPEPGLGGDFETENLRQCYQYCPV 50
Qy      40 -----GRECDCAQ----- 57
Db      51 LADIECRAAQLPDMLEBLGQQVDCDRMGLMCANSQQSPPLCHDYELRVLCCEYVPCGP 110
Qy      58 ----- 57
Db      111 SPAGTSPQSLASTEPAVPTPTQTATETKTLWVTPSIRSTAALTSGSGSGPVTVT 170
Qy      58 -----DY-----ESF 62
Db      171 PSAPGTTTCQPCQWTEWFDSDYPKSEQLGQGVDSYDKIRAAAGHLCOQPKDIECQAEF 230
Qy      63 -----CAEVH----- 67
Db      231 PNWTLAQGVQKVDHFLGVCNRWQGVFMKCNVYRIVLCCSDHCHGRATTPPPTT 290
Qy      68 -----NPTSPSSKAPPS-----GA 84
```

```
Db      291 ELETATTTTQALFSTPQPTSSPGLTRAPPASTTAVPTLSEGLTSRVTSLGTATTGGP 350
Qy      85 SOTIKST-----TKRSPKPPNKKTKKVVIE----- 109
Db      351 ROSAGSTBTVGVATSLTPRSALPGTTSGLTWRPQQPPTLAPTATMATSRARPTGTAS 410
Qy      110 ---SEE-----ITEHVSVENQES----- 125
Db      411 TASKPLATSLAPTTLTSELSTQABTSTPRTETMTSPLNTTTSQGTTRCQPKCEWTF 470
Qy      126 ----- 125
Db      471 DVDFTSGVASGDMETFFENIRAAAGKMCWAPKASIECRAENYPEVSDQGVQLTCSLETG 530
Qy      126 -----SSSSSSSSSSSTTWIKIKSSKNSA 148
Db      531 LTCKNEDQTRFNMCNVRVLCDDYSHCPSTLATSTATPSSTPGTTMIL----- 583
Qy      149 ANRELQKLVKDKNKKRTKPKPPVDEAGSLDNGDFKVT--TPDTSTTQHNKVS 206
Db      584 -----TKPTTATTAGSTATASSTQATAGPHVSTI-----A 618
Qy      207 TSPKIITAK-----PINRPSLPNPSDTSKETS----- 235
Db      619 TTPVTSSKATPPSPGATAPALRSTATTPTATSTAIPTSSSLGTTWRLSQTTTMA 678
Qy      236 -----TVNKEITVETKETT-----TNKQSTDGKEKTTSAKETQSIKTSAK 278
Db      679 TMTATPSSTPETVHTSTVLTTATTATGATGVSATPSSTPGTAHTKVLTTTGTGTAATP 738
Qy      279 DLAP-----TSKVL-----AKETPKAE 295
Db      739 SSSPGEARLPVWISITTTTPTTRGSTVTPSSIPGTHPTVLTITTTTATGSMATPSS 798
Qy      296 TTTKG--PALTTPKPEPTPTT-----PKPASTTPKEPTPTTIKAP----- 334
Db      799 TOTSGTPPSLTT-----TATTITATGSTNPSSTPGTTPPIPPVLTATTATPAATSTVTPS 854
Qy      335 ---TTPKEPAPTTS-----APTTP-----KEPAPTITKEPA 364
Db      855 SALGTHITPPPVNTTATTHGRSLSSPSHTVCTANTSATSGILGTHITEPSTGSHTPA 914
Qy      365 PTT-----PKSPAPTTTKEPAPTITTKSAPT-----PK 392
Db      915 AITGTTQHSTPALSPSPSSRTTSPSPGTTTTPGHTTATRTTATATATPSKTRTSTLLPS 974
Qy      393 EP--APT----- 398
Db      975 QPSTAPITTVTMGCEPQCAMSEWLDYPMGPGSGGDFDYSNIRAAGGAVCEQPLGLE 1034
Qy      399 -----PKKP-----APTTPK 408
Db      1035 CRAAQPGVLELRELQGVCELDGLVCRNREQVKFCMCFNYEIRVFCNKGHCFSTPA 1094
Qy      409 EPAPTTPKE-----PTPTTPKEPAP-----TTKEPAPT----- 436
Db      1095 TSTATPSSTPGTWTILTELTTTATTTESTGSTATPTSTLRTAPPKVLTTATTPTVTS 1154
Qy      437 ---TPKEPAPTAPKKA-----PTTPK-----EPAP-----TTKEPAPTITTK 471
Db      1155 SKATPSSSGTATAPALRSTATTATSTATSVIPISSSLGTTWRLSQTTTPTATMSTATP 1214
Qy      472 BPSP-----TTPKEPAPTITTKSAPTITTKPAPT-----KSAPTTP 508
Db      1215 SSTPETAHTSTVLTAITATTGATGVSATPSSTPGTAHTTKVPTTTTGTATPSSSPGTA 1274
Qy      509 KSP-----SPT----- 515
Db      1275 LTFPPVWISITTTPTTRGSTVTPSSIPGTHATVLTITTTTATGSMATPSSSTQTSCTP 1334
Qy      516 -----TKSPAPTTPKEPAP-----TTPKKAPPT-----TPKEPAPTTPKEPA 552
Db      1335 PSLTTTATTATGTTNPSSTPGTRPPIPPVLTATTATTAATSTSTVTPSSALGTHHTPPV 1394
```


QY 553 PTTT----- 556
DB 1395 PNTTATTHGRSLSPSPSEHVTWTAWSATSGTLGTHITBPSGTGHTPAATGTTQHSTP 1454
QY 557 ---KKPAPT---APKEPAPTTP---KETAPTPKK---LTPTPEKLAPTTP- 596
DB 1455 ALSSPHPSRTTESPPSGTTTPGHTTATSRITATATSKTRTSTLLPSSPTS-APITTV 1513
QY 597 -----EKP----- 599
DB 1514 VTMGCBQCAWSEWLDYSYMPGPGGGDFDTYSNIRAAGGAVCEQPLGLECSAQAGVP 1573
QY 600 -----APTPEELAPTPEE- 614
DB 1574 LRELQGVVECSLDFGLVCRNREQVKMCFNVEIRVFCNNGHCPSPATSTATPSST 1633
QY 615 -----PTPTTPEPAP-----TTP-----KAAAPNTPK 637
DB 1634 PGTTWILTQTTAAATTTATTGTAIPSSSTPGTAPPKVLTSQATTPATSSKATSSSPR 1693
QY 638 -----EPAPT-----PKEP----- 647
DB 1694 TATLVLSTATKSTATSTFTIPSSLTGTTGTSQNRPPHPMATSTIHPSSTPTHTS 1753
QY 648 -----APTTKPEPAPTTPKETAPT-----PKGTAPT-----LKEP 679
DB 1754 TVLTKATTTTRATSSMSTSPGTTWILTAAATTAALPHGPPSSTPGTTWILTEP 1813
QY 680 APT-----TPKK-----PAPKE 691
DB 1814 STTATVTPGTSTASTRATAGLKVLTSTATTPTVISSRATPSSPGTATAPALRS 1873
QY 692 LAPT----- 695
DB 1874 TATPTATSVTAIPSSLSGTATWRLSQTTTPTATMTATPSTPTETVHTSTVLTTATTT 1933
QY 696 -----TTKEPTSTT-----SDKPAPT----- 712
DB 1934 RTGSAVATPSTPGTAHTTKVPTTTTGTATPSSSPGTALTPPVWISITTTTTRGSTV 1993
QY 713 -----PKG 715
DB 1994 PSSIPGTHATVLTITTTTAVATGSMATPSSSTQTSCTGTPPSLTATTATGSTTNPS 2053
QY 716 TAPTTKPEPAPT-----PKEP--AP 734
DB 2054 TPGTTPIPPVLITATTAPATSTVTPSSALGTHTPPVVENTATTGHSLPSPSPHTVP 2113
QY 735 T-----TPKGTAPT-----LKEPAP-----TPKKAPKEL 761
DB 2114 TAWTSATSGILGTHITEPSTGTSHTPAATGTTQSTPALSSPHPSRTTESPPSP--- 2170
QY 762 APTTTKGPTSTSDKPAPTTPKET-----AP-----TTPKEP----- 793
DB 2171 -GTTTPTGTRGTSRTTATATPSKTRITSLPSPSAPITTVTTGCEPQCAWSEWLDYS 2229
QY 794 ----- 793
DB 2230 YPMGPGSGGDFDTYSNIRAAGGAVCEQPLGLECRQAQGVPLRELQGVVECSLDFGLVC 2289
QY 794 -----APTTPKAPPTTPETPPPT----- 813
DB 2290 RNREQVKMCFNVEIRVFCNNGHCPSPATSTATSPSTPGTTWILTKLTTATTTE 2349
QY 814 -----SEVSTPTTT 822
DB 2350 STGSTATPSSGPGAGTPHVSATTATPTVTSSKATPSSPGTATAPALPALRSTATTPTAT 2409
QY 823 -----KEPTTIKS----- 831
DB 2410 SFTAIPSSSLGTTWRLSQTTTPTMATMTATPSTPTETVHTSTVLTTATTATTGATSVAT 2469

QY 832 ----- 831
DB 2470 PSSPPTGTAHTTKVPTTTTGTFTVTPSSSPGATPVPWISTTTTPTTSGSTVTPSSIPGT 2529
QY 832 -----PDEST----- 836
DB 2530 THFTVLTTTTPQVATGSMATPSSSTQTSCTGTPSLITATTATGTTNPSSPTPGTPI 2589
QY 837 -PELSAEPPTKALEN-----SPKEP----- 855
DB 2590 PPELTATTATPAATSTVTPSSALGTHTPPVNNTATTGHSLSLSPSPHTVTAMTSAT 2649
QY 856 -----GVPTTKTTPAATPEMT-----TTA 874
DB 2650 SGTLGTHITEPSTGTSHTPAATGTTTSTPALSSPHPSRTTESPPSGTTTPGHTTA 2709
QY 875 KDKTTE----- 880
DB 2710 TSRTTATATPSKTRITSLPSPSTAPITTVTTGCBQCAWSEWLDYSYMPGPGGDF 2769
QY 881 ----- 880
DB 2770 DYSNIRAAGGAVCEQPLGLECRATAQGVPLGELQGVVECSLDFGLVCRNREQVKFKM 2829
QY 881 -----RDLRTTPETT-----TAAPKMT 897
DB 2830 CPNVEIRVFCNNGHCPSPATSTATSPSTPGTTWILTTELTTATTASTGSTATPSST 2889
QY 898 KETA-----TTTEKTTESKI-----TATTTCOV-TSTTTQDT-----TPEKITLTK 936
DB 2890 PGTAAPPKVLTSPTATTPTATSSKATSSSPRTATLPLVLTATKSTATSVTPIPSSTLG 2949
QY 937 TTTTLAPKVVTTK-KTITTTTMMKPEE-----TAKPKDRATNKSAT----- 976
DB 2950 TTGTLPSQTTTPVATMTSTIHPSSTPTETHTSTVLTTKATTRATSTSTPSSPTGTTWILT 3009
QY 977 -----TPKPQK----- 982
DB 3010 ELTTAATTTAGTGPTATPSSPTPGTTWILTTELTTATTASTGSTATLSTPGTTWILTEP 3069
QY 983 -----PTKAPKPTSTKKPKT---MPVRKRP 1005
DB 3070 STTATVTPGSTATASTAQTAGTPHVSATTATPTVTSSKATPSSPGTATAPALRST 3129
QY 1006 KTTTTPPKMTSTMBELNPTS-----RIAAEMLQTTTRNQTNQNSKLVEVNPKEADG- 1057
DB 3130 ATTTPTATSTAI-----PSSSLGTTWRLSQTTTPTATMTSTATPSS-----TPETVHTST 3179
QY 1058 -----GAGGSTPHMLLRPHVMEVTPDMVDLPRVP-----NQGIILN 1095
DB 3180 VLTTTATTGATGS-----VATPSTPGTAHTTKVPTTTTGTATPSSSPGTAIT 3230
QY 1096 P 1096
DB 3231 P 3231

RESULT 8
S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N.Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein VIRO19C
C.Species: Saccharomyces cerevisiae
C.Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004
C.Accession: S48478; A26877; B26877; S27281; JC6123
R.Rowley, K.
submitted to the EMBL Data Library, October 1994
A.Reference number: S48478
A.Accession: S48478
A.Molecule type: DNA
A.Residues: 1-1367 <ROW>
A.Cross-references: UNIPROT:P08640; GB:Z47047; EMBL:Z38061; NID:G603997; PID:G763364; GS
R.Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987

A:Title: Gene fusion is a possible mechanism underlying the evolution of STA1.

A:Reference number: A91831; MUID:87194600; PMID:3106330

A:Accession: A28677

A:Molecule type: DNA

A:Residues: 1-242 <YAM>

A:Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525

A:Accession: B26877

A:Molecule type: DNA

A:Residues: 762-1331 <YA2>

A:Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526

R:Parado, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.

FBS Lett. 239, 179-184, 1988

A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from *Saccharomyces cerevisiae*

A:Reference number: S27281; MUID:89031230; PMID:3141213

A:Accession: S27281

A:Molecule type: DNA

A:Residues: 1-31 <PAR>

A:Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552

R:Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.

Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996

A:Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohyphae formation in *Saccharomyces cerevisiae*

A:Reference number: JG6123; MUID:96323237; PMID:9710886

A:Accession: JG6123

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-1367 <LAM>

A:Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387

C:Genetics:

A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458

A:Cross-references: MIPS:Y1R019c; SGD:S0001458

A:Map position: 9R

C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

C:Keywords: glucosidase; hydrolase; polysaccharide degradation; transmembrane protein

F:5-21/Domain: transmembrane #status predicted <TM1>

F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 16.0%; Score 933.8; DB 1; Length 1367;

Best Local Similarity 25.6%; Pred. No. 1.6e-11;

Matches 350; Conservative 110; Mismatches 446; Indels 462; Gaps 69;

QY 9 YLL-----LILLVVFVQVSSQBLSCGRC-----CQVMPNFQCFEYLGSSAAQYASSQWG 169

DB 118 YLLDNPDTDFATFVATQDVS-----LILLVVFVQVSSQBLSCGRC-----FE----- 35

QY 36 --SPERGECDDAOCKYKCCPDYEFCAEVHNPSTPPSSKAPPPSGASQT----- 87

DB 170 TTSF-----DLSTGCNNYN-----QHSQTDFFGFY 196

QY 88 -----IKSTTKSPKPPNKKTKKVVSEBEITEHSVENQSSSSSSSSSSSS 136

DB 197 WNIIDCNCGGTKSSITTSSESTTSSTSESTTSSTSESTTSSTSESTTSSTSESTTSST 256

QY 137 TIWKIKSKNSAANRELOKKLVKDNKKNRKKKPT-----PKPPVVDGAGSL 185

DB 257 TAPATPTTSC-----TKRKPTPTTSCCKRKPTPHHD----- 291

QY 186 DNGDFKVTTPDSTQHNKVSFSPKITTAKPINRP-----SLP-----PNSDTSKE 232

DB 292 -----TTPCT-----KKTTTSKTCYKKTTPVPTSSSTTSSSAPVTPSSSTTES 339

QY 233 TSLVYNKETT-----VETKETTNNKQTS-----TDGKTKTSKAKTQSIKTSKXDLAPTS 284

DB 340 SSAPVTSSTTSSSAPVTPSSSTTSSSAPVTSSTTSSSAPVTSSTTSSSAPVTSSTTSSS 395

QY 285 KVLAKPTKAEVTTKGA-----LTPKPTPTTPKPASTTPKPTPT-----IKSAP- 334

DB 396 -----PTSSSTTSSSAPVTSSTTSSSAPV-----SSSTSSSAPVTSSTTSSSAPV 446

QY 335 ---TTPKPAPTTKSAPTTPKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTP 391

DB 447 TSSTTSSSAPVTPSSSTTSSSAPVTS-----TSSSAPVTPSSSTTSSSAPVTP- 501

QY 392 KEPATTPKKAP-----TTPKPAPTTKKPTP-----TTPKPAPTTKKPAPTTKKPA 441

DB 502 ---SSITSSSAPVTPSSSTTSSSAPA-----PTPSSSTTSSSAPVT---SSITSSS 551

QY 442 APTAPKKPAPTTPKPAPTTP-----KPAPTTKKPSPTTPKPAPTTKKPAPTTKKPAPTT 492

DB 552 APV-----PTPSSSTTSSSAPVTPSSSTTSSSAPVTPS---SSITSSSAPVTPSSSTT 604

QY 493 TKEPAPTTKSAPTTPKPSPTTPKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKK 551

DB 605 ESSSAPAPTSSSTTSSSAPVTS-----TSSSAPVTPSSSTTSSSAPVTP-----P 656

QY 552 APTTKKPAPTAPKPAPTTPKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTT 611

DB 657 SSSTTSSSAPVTPSSSTTSSSAPVT---SSITSSSAPVT---SSITSSSAPV- 707

QY 612 PEPPTP---TTPKEPAPT-TPKAAAPNTKPAAP-----TTPKEPAPTTPKPAPTTKPE 662

DB 708 ---PTPSSSTTSSSAPVTPSSSTTSSSAPVTPSSSTTSSSAPVT---SSITSSS 760

QY 663 TAPT-TPKGTAPTTLKPAPTTPKPAKXELAPTTPKPTSTSDK---PAPTTPKGTAP 718

DB 761 SAPVTPSSSTTSSSAPVTPSSSTTSSSAPVT--PSSSTTSSSAPVP-TPSSSN 817

QY 719 TTPKEAPT-----TPKEPAPTTPKGTAP-----TTLKPAPTTPKPAKXELAP 762

DB 818 ITSSAPSTPSSSTTSSSAPVTPSSSTTSSSAPVTPSSSTTSSSAPV-----PTPSSS 873

QY 763 PTTTKGP-----TSSTSDKPAPTTPKPAKXELAPTTPKPAPTTPKPAPTTPET---PPTTSE 815

DB 874 NITSSAPSSITPSSSTTSSSTGTT---VTPSSSKYPSQTSVSTSTTIVPKTITS 930

QY 816 VSTPTTKKPTTHKSPDSTPELGAETPKALENSPKBPQVPTTKTTPAATKPEMTT--- 872

DB 931 VTPSTTTTITVTCSTGTNSAGETTSKSPKVTIT---VPTTTTTSVTSSTTTT 985

QY 873 ---TAKDXTBERDLRTPTT-TTAAKPKTKTATTTEKTTESKITATTTOVTS- -TTO 925

DB 986 TVCSTGTNSAGETTSKSPKTTTTPVPCSTSESTASESTTTPPTVTVTVTVTVTVITE 1045

QY 926 DTPPFK-----ITLTKTTTLA--PKVTT---TKKTTTTEI---MNKPEET 963

DB 1046 YSTSTKPGGEITFTFVKNIPTTLTTLAPTSEVITVNTPTTTTITVTCSTGTNSAGET 1105

QY 964 ---AKPK-----DRATNS--KATN----- 977

DB 1106 TSCSPKTVTVTVPCSTGTGEVTEATTLVTAVTTVTVTVTVTVTVTVTVTVTVTVTVTK 1165

QY 978 --PKPKPKTKAPKKP-----TSTKPKPTM----- 999

DB 1166 SVPTTV 1225

QY 1000 ---PRVKPKTTTPPKWSTMPELN-----PTSRIEAM----- 1031

DB 1226 TAPSATPTVTAIPTVITTESVGTNSAGETTTGYTKSIPTTITLIPGNSGAKNYE 1285

QY 1032 -LQTTTRPQNTNSKLVEVNPKESEDAGGAEGETPHMLLRPHVFMPEVT 1078

DB 1286 TVATAINP-----ISKTTSQLATTASAS---VAPVTVTSPLT 1321

RESULT 9

T34513

hypothetical protein ZK783.1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004

C:Accession: T34513

R:Favell, A.; Vaudin, M.

A:Submitted to the EMBL Data Library, August 1994

A:Description: The sequence of *C. elegans* cosmid ZK783.

A:Reference number: Z21536

A:Accession: T34513

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3507 <FAV>	QY	400	-----KKPAPTTTPKEPAPTT-----	414
A:Cross-references: UNIPROT:Q23587; EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783	Db	1378	PQNRNKEPEPTKDTFALPTTTTGAPQANDSVVENTKCTSSDEGLDLCALCERRTGVCRCPEP	1437
A:Experimental source: strain Bristol N2; clone ZK783	QY	415	-----PKE-----	419
A:Gene: CESP:ZK783.1	Db	1438	GFEGAPPKSKCVDVDECATGDHNCHEARCONVYGVACFCPTGFRKADGSCQDIDECT	1497
A:Map position: 3	QY	420	-----PTTPKEPAPTTK-----	431
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2; 3504/1	Db	1498	EHNSTCCGANAKVKNKPGTYSCCEENGFLGDGYCQVPTT-KKPCDSTQSSKSHCSBNMS	1556
Query Match 15.4%; Score 896.2; DB 2; Length 3507;	QY	432	-----EPAP-----	435
Best Local Similarity 14.9%; Pred. No. 3.4e-10;	Db	1557	CEVTVDSVCKEOMGGYKKGKVCEDINECVAEKAPCSLNANCVMNMNAGTTFSCCKQGY	1616
Matches 386; Conservative 177; Mismatches 444; Indels 1578; Gaps 84;	QY	436	-----TTTPE-----	440
QY 19 IQOVSSQELSCGRCFESPE-RGRBCDDA-----CKKYDKC-----	Db	1617	RGDFMCTDINECDERHPCHPBAECTNLEGSFKCECHSGFEGDGKIKCTNPLERSCEDVE	1676
Db 506 VELTSGRLACTSYCPNPSECYGCYCEVSGYGGNALVGCEDIDCITEICNIEANECV	QY	441	-----PA-----	442
QY 56 -----CPDY-----	Db	1677	KFCGRVDHVSCLSVRIYNGSLSSVCECEPOFRPEKESNCVDIDECESRNNCDPASAVC	1736
Db 566 NLIIGFVCCNPNTATHDDCIDFLTKVIYAYMIIFLLKGLKITKEKGLHVLVIGNEEDTV	QY	443	-----	442
QY 60 -----ESFCAEVH-----	Db	1737	VNTGSRCECAEYEGEGVCTDIDECRGMAGCDSMAMCINRMGSCGCKMAGYTGCG	1796
Db 626 VATRSNHSSTDLITQVQSRNFSSTGQIILTRGVSGEAVTQTDADEFLGLEISAADL	QY	443	-----KPAPTTTPKEPAP-----	459
QY 68 -----NPTSPSSKKAPPPSGA-----SOTIKST--	Db	1797	ATCIKIBEPKSDKTACTDEWSRLCELEKKQCTVDEBEVPQCGACLPGHHPINGTCQSLQ	1856
Db 686 AGSGGITLPTLEPKIEGSKKA---SGVWTEDEGEDEDLMEEGSGSWTTINGTGI	QY	460	-----	459
QY 92 -----TKRSPPPKKKTKKVIES---EEITEEHSVSEN	Db	1857	ISGLCAQKNDCKNAECIDIHPSHFHFCSPDGFIDGMI CDDVDECNAGMCDDENTKCE	1916
Db 743 TGSPRSEGTIRVITLGEDGPEPATKPGISAPDKTGEKSTESDGEKLTVEKDGEA	QY	460	-----TTTPEP-----	484
QY 123 QESSSSSSSS	Db	1917	NTIGSFNCVCLGFKFYDEKCVVDEKKQPNREKIEIDEENSSSSSQEKPTTKGIVSST	1976
Db 803 QSSGSSATSGKXKSEATSGSSSSSAKSGTGSEASGSSGASSSGSGVSGSSSVTSESG	QY	485	TTKSAPTTTTPKEPAPTTTTSKAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEP	544
QY 133 -----SSSTTIWKIKSSKN-----	Db	1977	SATSSESTTAEPHVTTISSTTSTKDMTSSKSPENVTMSSESPEVSTSSSKSTTASSETTV	2036
Db 863 SFGTSSSGVGEATSGTVDGSESGKPSKSTEEKLPFTKNGEKSPISGSDTTGKESSE	QY	545	PTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEP	595
QY 147 -----SAANRELQKLKV-----	Db	2037	SSTPSESS-----SSEAPLTSSPATTEV--ITESSVKSTTPKESSSEITVKLSKS	2087
Db 923 ETTSRPIEGSDSLTEGSGGEWFETGSKGHFESGKSVSTGKGTQSGAEGSGSGPKV	QY	596	PE-----KPAPTTPEELAPTTPEEPPTTPEEPAPTTTPEEPAPTTTPEEPAPTTTPEEPAPTTTPEEPAP	643
QY 160 -----KDNK-----	Db	2088	PEVTSSVKSSPSTPS-----TTSQSVTSTVETSKSTVLSSEAPVTSTSTEVHTVSETK	2143
Db 983 PKGPGAPEITTDGEESSTSTGKSGGPKADKSDKNVKTGDKNPDITTDGEDSTSETS	QY	644	PKEPAPTTTPKEPAPTTTPKET-----APTTPKGT-----APTTLKEPAPTTTPKPAKPELA	693
QY 164 -----KNRTK-----	Db	2144	PSLSASSITGDTNSTTTPSTSLASVKSTSAPEGTSASVAPVKLSLSPDV-SQPSTKTFD	2202
Db 1043 GGEQGPKGKQPPGDKGSEVKKPTSEVDGPNLSTGKSKNVPLKPTDLPPEGSGI--	QY	694	PT-----TTKEP-----TSTTSDKPAPTTTPKGTAPTTTPKEPAPTTTPK	730
QY 188 GDFKVTTPD---TSTTQH-----	Db	2203	ATESSTVQASSTSGTSVKSTSEPESHVTKLSITSSNPSSSVPTSPKSTPTVPEST--E	2260
Db 1101 -----LTSSGGKNSTFEHGTLERLPPKTEDKSSETPQLGLEISAKKPE--PEDGTSKE	QY	731	EPAPTTPKG-----TAPTTTKEPAPTT-TPKKP	756
QY 233 TSLTVNKETTVETKETTTNKOT-----	Db	2261	QPTSTTPGQSLTPMNSNSEVLTTSEPHVLSLSPDVSSQSTTPNNLSSTTETPKTS	2320
Db 1155 VGLLEIWESE--TPGSTLLSDSVGLEISGSLTKATKXPHVEIEGSGTGDEIETAT--T	QY	757	APKEL---APTTPKGPT-----STT-----SDKP-----	777
QY 270 QSLKETSATK-----DLAPTSKVLAKPTPKAETTTKGPALTTPKEPT-----	Db	2321	SEVLSNEEPSTTEAPTTLSPDILSTTNNLSQSSVTSTEDRSRISSENSEKPTSAPELV	2380
Db 1211 RDVSKSTKPRVVDGDNGETSGVDGKPTTAPTPSSSAESSTGERIPTTSASPEGSGG	QY	778	-----APTTP-----KETAPTTT-----KEPAPTT	797
QY 311 -----PTTPK---EPASTTKEPTPTTIKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP	Db	2381	TSSVTHVASSPDVPTESSEPDLLTGSTENIPEASSKQTISSPTPTDPTTASSEPTKST	2440
Db 1271 EAGVPSPDGSGESTSAPDGVSPSTATAPEVP-----TTSASSITPDVAEEGIGPST	QY	798	PKKP-----APTTP-----	807
QY 360 TK---BPAPTTTKEPAPTT-----TKPAPTTTTSKAPTTTKEPAPTTTKEPAPTTTKEPAP	Db			
Db 1324 SKPTAEPLETT---APSTEVTSPEGSGTEESTLPPTEGSGESTTSSAPTIV--EPATVLP	QY			

T29757		protein UNC-89 - Caenorhabditis elegans	
C:Species: Caenorhabditis elegans			
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999			
C:Accession: T29757			
R:Du, Z.; Le, T.T.; Wilson, R.			
submitted to the EMBL Data Library, May 1997			
A:Description: The sequence of C. elegans cosmid C09D1.			
A:Reference number: Z20679			
A:Accession: T29757			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-6642 <DUZ>			
A:Cross-references: EMBL:AF003131; PIDN:AAE54132.1; GSPDB:GN00019; CBSP:unc-89			
A:Experimental source: strain Bristol N2; clone C09D1			
C:Genetics:			
A:Gene: CBSP:unc-89			
A:Map position: 1			
A:Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6			
/3; 591/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1			
Query Match		15.1%; Score 879.9; DB 2; Length 6642;	
Best Local Similarity		17.1%; Pred. No. 1.8e-09;	
Matches 379; Conservative 173; Mismatches 419; Indels 1251; Gaps 89;			
QY	9	YLLLLSVFVIOQVSSQE-----LSCKGRCFESFERGREGC---DCDAQCKK 51	
DB	393	YSIRLDKYNIRQHTTDETVLQPOEGLPSFRPKDFETSEYVRKAWLEDIAEEQEK 452	
QY	52	Y-----DKCCPDYEFSCARVHNPT-----SPSSK-K 77	
DB	453	YAAERDAISMTASEMTASSVDFMNASDQSEFSEWSGRKSSLFPGPEGGPRKKVK 512	
QY	78	APP---PSGASOTIKS-----TTKRS-----95	
DB	513	SPVVISPTGSSISVSGSSSIDWTTGTTLEMQGTRVTRTOYGRFTQESSAKNCLKVT 572	
QY	96	--PKPPNKKTKKVISEBIT-----EHSVSENQBS-----125	
DB	573	GYPPLP-----DITWKDDVQLHEDERTFYSDGDFAMTIDPVQVDTGKY 619	
QY	126	-----SSSSSSSSSTTWIKSSKNSAA-----NRELQK-----155	
DB	620	TCMATNEYGOASTSAFFRVLKVEKEAAPFAFVKURDKCEKGDVIDFECEVGEWPEPEL 679	
QY	156	-----KLKVKDNK-----163	
DB	680	VMLVDQPLRPHDFLQYDGTAKLEIRDAQPDGTVVTVKIQNEFGSIESKAELFVQA 739	
QY	164	---KN-----165	
DB	740	DPDKXNHVAFQATIEYVCEDEGEYRFRKSVITGDPNPPIIWFINGKPLSESEKVFISE 799	
QY	166	-----RTKKKPTPKPPVVD-----180	
DB	800	DGICILTIKDVTHFDGWTGCGSNELGSASCDGRKLVAVPAPPTFNKPLSDKTVQEK 859	
QY	181	-----AGGLDNGD-----FKVTPDPTSTTQHN-----203	
DB	860	TUVFEVDVSGWPPTLTFTLCKGKELXNGEEVIGHDGFYRISIPNTSMXKHGDIKAV 919	
QY	204	-----KVSTSPKLTAKPI-----NPRSL 223	
DB	920	AQNEHGTASRLKTVQEEESRSAP--TFLKIDEDQVTKGFAVFETTVRGNPDEV 977	
QY	224	-----223	
DB	978	TWFINGHKMDQSGPGVKIEAHNHDHKLITIDSAQYAGTVLCRAENAVGRFETKARLVVLAP 1037	
QY	224	-----PPNSDTSKETSILTNNKETT-----ETKEITT-----250	
DB	1038	EKQKXPP-----KFVEILVDKTEVDNTVVFEVRVEGEKPTVTWYLKGEELKQSDRVEI 1092	

QY	251	-----TNKQTSDDGKEKTTSAKETQSIEKTSKADLAPTS 284	
DB	1093	REFDGSIKISIKIKIEDAGEIRAVATNSEGSDTKAKLTQVKRPAPF-----FDRPVS 1148	
QY	285	KVLAK-----PTP-----KAPTITKGPALT-----305	
DB	1149	LTVEKGEAVFSAHAFGIPLPTVEMVNGRKVRDQEGARVTRDESVDGASILTIDTAT 1208	
QY	306	-----PKBPTTTPKEPAST-----T 321	
DB	1209	YSEVNHLLTISVAENLTGABETGAQLTIEPKKESVVVEKQDLSSEVQKEIAQOVKEAS 1268	
QY	322	PKEPTPTTIKSAPTTTPKEPAPTT-----KSAPT-----350	
DB	1269	PEATTITWETSILTSKTTMTSTTEVTSTGVTVETKESESESATTVIGGSGGVTEGS 1328	
QY	351	-----TPKEPAPTTTKEPAP-----TTPKE-----PAPTTTKEPAPT 382	
DB	1329	ISVSKIEWKTSQSDQVREGTRRRVSPAEBELPKEVIDSDRKKKSPSP--DKKEKSP 1387	
QY	383	TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP--KE 440	
DB	1388	KTEKPAFTPKTGEVVKSP---KESPAFTPKKEKSPAABEVKSPTKKESPSPTKE 1444	
QY	441	PAPTAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKPSPTTP-----KEPAP-----483	
DB	1445	KSPSPKTKTGDEVKEKSP--PKSP---TKKESPEKPDVKSPVKKEKSPDATNIVEVS 1499	
QY	484	-----TKBAPTITKSAPTTP-----KEPSPTTTPKEPAP 521	
DB	1500	SETTIEKTTMTTMTHESEBSRSTVKKEKTEPEKVDKPKSPTKKDKSEKSIIEIKS 1559	
QY	522	TTPKEPAP--TTPKKPAPTTTPKEPAPTTTPKEPAPTTTK-----PAPT-----562	
DB	1560	PVKKEKSPKVEEPKSPATKKEKSPKSPASP---TKKSENEVKSPATKKEKSPKSVVEE 1615	
QY	563	--APKEPAPTTPEATPTPKLTPTTPEKLAFTTPEKPA-----PTTPEELAPTTPEE 614	
DB	1616	LKSPKESKSP---EKADDKPK--SPTKKEK---SPEKATEDVKSPATKKEKSPKVEEK 1665	
QY	615	PTPTTPEEAPPT--TPKAAAPNTPEKAPTTPKPE--PAPTTTPKEPAP-----TTPKET 663	
DB	1666	PTSPATKESSTPKTDDVKSPATKKEKSPQVVEKPKSPATKKEKSPKSVVEEKSPKEK 1725	
QY	664	APTTPKGTAPTTLKEPAPTTTPKAPKAPKELAPTTT-----KETSTTSDKAPPT 711	
DB	1726	SP-----EKABEKPSPATKKEKSPKSAABEVKSPATKKEKSPKSAABEKPSP 1773	
QY	712	TPKGTAPT-----TPKEPAPTTTPKEPAPTTTPKGTAPTTLKEPAPTTTPKAPKAPKEL 761	
DB	1774	TKKESPVKXADDEVKSPATKKEKSPKVEE--KPAKP-----TKKE--KTPEKSAABEL 1823	
QY	762	APTPTTK-----GPTSTTSD-----KPAPTTTPKETAPTTTPKEPAPTTTPKAPPT 804	
DB	1824	KSPTKKEKSPSPATKKTGDSKESKPEKEPKSPATPKSPGSPKPKKSKSPKAPKPP 1883	
QY	805	TPETPTPTTSEVSTPT-----TTKEPTTTHKSPD---833	
DB	1884	APK-----LTRDLKLQTVNTDLAHFEVVVHATECKMFLDCKEITTAQGVTVSKDDQFEP 1939	
QY	834	-----ESTPELSAEPKPALENSPKPEGVPTTKTFAATKP 868	
DB	1940	RCSIDTTWFGSGTVSVVNAAGSVETKELKLETPK-----ETK-----KP 1982	
QY	869	EMTTAKD-----KTTED-----882	
DB	1983	EFTDKLRDMEVTKGDTVQMDVIALHSPLYKQWYQNGNLELDGKNGVTKNENKSSLIIPN 2042	
QY	883	-----LRTTPTETTTA---APKMT---KETATTTETKTES---910	
DB	2043	AQDSGKITVEASNEVGSSSSAQLTVNPPSTPIVVDGPKSVTIKETETAEFKATISGPP 2102	
QY	911	-----KITATTTQVTSITTTQDTPPKIT-----TLKTT-----938	

```

Db      2103 APTVKKTINEKIVBESRTTITIKTEDVYTLKISNAKIBOTGTVKVTAQNSAGDSQADL 2162
QY      939 -----TLAPKVTITTKTIT-----TTE 955
Db      2163 KVEPNVKAPE---KSLQIDKVADEGEPLRWNLDELDPSPGTEVSWLLNGQPLTKSDIVQ 2219
QY      956 IMN-----KPEE---TAKPKDRA---TNSKATTPKQPTKAPKPTS--- 992
Db      2220 VDHGSGTYHTIAEAKPEMSGTLTAKAKAAGECETSATKVTWNGNKKPEFVQAPQNH 2279
QY      993 -----TKKP-----KTMPEVRKP----- 1005
Db      2280 TTLEESVKFSAITGKMPNVTWYLNKKLQSEBVKVYHETGKTSIRIQKPLMEHG 2339
QY      1006 -----KTTPTPRKMTSTM-----PE--- 1020
Db      2340 TIRVEAENVGKVOATAQAKVDKTEVP-KFTTNMDRQVKEGEDVKFTANVEGYPEPSV 2398
QY      1021 ---LN---PTSRJAEAMLQTTTRPNOTPSK---LVEVNPKSEDAGG----- 1058
Db      2399 ANTLNGEVPK-----HPNITVTDKGEHTIEISAVTPEQAGELSCAETNPFVGSK 2448
QY      1059 -----AEGETPHMLLRPHVFM-----EVTDM 1081
Db      2449 KRQVQLAVKVGDAPTFAKNLEDRLITEGELTLMDAKNLNIVKPKKITWLKDGVEITS 2508
QY      1082 DY 1083
Db      2509 HY 2510

RESULT 12
B48666
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 15-Mar-2004
C:Accession: B48666
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq
ins.
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: B48666
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2897 <SCH>
A:Cross-references: EMBL:X65551
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F:29-91/Domain: kinase interaction domain homology <KIH>

Query Match      15.0%; Score 875.8; DB 2; Length 2897;
Best Local Similarity 14.9%; Pred. No. 6.5e-10;
Matches 427; Conservative 138; Mismatches 439; Indels 1862; Gaps 104;

QY      3 WKT-----LPIYLLLLSVFVIQVSSQELSCRCFESPERGECDCDA 47
Db      2 WPTRLVTIKRSGVDPHFPLSLTCL-----FGGIECDIRI 39
QY      48 Q-----CKYDKCCPDYESFCAEVH-----NPT----- 70
Db      40 QLPVVSKQHCK-----IEIHEQEAHLNFSNTPTQVNGSVIDEVRLKHGD 86
QY      71 -----SPPSSKKAPPPSGASQT 87
Db      87 VITIIDRSFRYENSLQNGRKSTEFPRKIREQEPARRVSRSSFSDDPEKAQDSKAYSKI 146
QY      88 IKSTTKRSFK-----PPN-----KKKT-----KKVIESEBITTEHSYSEN 122
Db      147 TEGKVSNGNPELFDENLPNTPLKRGAPTKEKSLVMHTPPVLKLIK-----EQQPSGK 201
QY      123 QESS-----SSSSSSSST----- 137

```

```

Db      202 QESGEIHVEVKAQSLVISPPAPSPKTPVASDQRRRSCKTAPASSSKSQTEVPKRGGER 261
QY      138 -----IWKIKSSNSAANRELQK-----KLKVKDNK----- 163
Db      262 VATCLOKRVISRSQHDILQMICSKRRSGASEANLIVAKSWADVVKLGAKQOTQTKVIKHG 321
QY      164 ---KRTKKKP---TPKPPV----- 177
Db      322 PORSMMKQRRPATPKPKPVEVHSQFSTGHANSPTIIIGKAHTEKVHVPAFYPYRLNFF 381
QY      178 ---VDEAG-----SGLDNGD----- 189
Db      382 ISNQKMDFKEDLSGIAEMFKTPVKEQPOLTSTCHIAISNSENLLGKQFOGTSGBEPLJP 441
QY      190 -----PKVT-----TPD 196
Db      442 TSESGFNGVFFSAQNAAKOPSKCSASPLRQCIRENGVNAKTPTNTYKMTSLETKTSD 501
QY      197 TSTTOHNVST---SPKITTAKPINRP----- 221
Db      502 TETEPSKTVSTVNRSGRSTEFNIOKLPIVESKSEETNTEIVECILKRGOKATLLOQRREG 561
QY      222 -----SLP----- 224
Db      562 EMKEIERPFPETYKENTELKENDEKMKAMKRSRTWGQKCAPMSDLTDLKSLPDTLMKDTA 621
QY      225 -----PNSDTSKETSLTVNK-----ETTVETK----- 246
Db      622 RQNLLOTQDHAKAPKSEKGIKMPQCOSLQPEPINTPTHTKOQLKASLGKVGVEELIA 681
QY      247 -----ETTTTNKQTSQDQKE-----KTTSAKETQSIE 273
Db      682 VGKFTRTSGETHTHREPPAGDGKSIITFKESPQIILDPAAVGTGMMKWPRTPKBAQSLIE 741
QY      274 KTSKDLAPTSKVLAKPTPKAE-----TTTK----- 299
Db      742 ---DLAGFKELFQTPGPSEESTDEKTTKIACKSPPEVDPTPTSTKQWPKRSLRKAD 796
QY      300 -----GPALTTPKPE-----TPT-----TPK 315
Db      797 VSEEFIALRLKLTSPSAGKAMLTTPKAGGDEKDIKAFWGTVPQKLDLAGLPGSKRQLOTPK 856
QY      316 EPAST-----TPKPEP-----TPTIKSAFT----- 335
Db      857 EKAQALEDLAGFKELFQTPGHTTEELVAAGKTKIPCDSPQSDPVDPTPTSTKQPKRSIRK 916
QY      336 -----TPKEPAP-----TTTSA 348
Db      917 ADVEGELLACRNLMPSAGKAMHTPK---PSVGEEDIIIFVGTVPQKLDLENLTGSKRR 973
QY      349 PTPKPEPAPT-----TTKEPAPTTPKPEPAPTTPKPEPAPT 383
Db      974 PTPKPEEAQALEDLTGFKELFQTPGHTTEELVAAGKTKIPCDSPQSDPVDPTPTSTKQPKRSIRK 1027
QY      384 TKSAPTTPKE-----PA 395
Db      1028 TRQPKTPLEKRDVOKELSAKLKLTQTSGETHTDKVPGGEDKSIINAFRETAKQLDPAA 1087
QY      396 PTPPKPAPTTPKEPA-----PTTPK-----EPTPT-TPK 424
Db      1088 SVTSGKRHEKT-KERAAQPLEDLAGWKLFPQVCTDKPTTHEKTTKIACRQSDPVDPTPT 1146
QY      425 EPAPTKE-----PAPT----- 436
Db      1147 SSKPOSKRSLRKVDVEEVEFFALRKRTSPSAGKAMHTPKPAVSGEKNIYAFMGTPVQKLDLT 1206
QY      437 -----TPKEPA-----PTAPKK 448
Db      1207 ENLTGSKRLQTPKEKAQALEDLAGFKELFQTRGHTTESMTNDKTAKVACKSSQPDLDKN 1266
QY      449 PA-----PTTPKEPAPT----- 460
Db      1267 PASKRRLKTSLGKVGKVEELLA VGLTQTSGETHTHTTEPTGDKSMKAFMESPQILD 1326

```

```
Qy 461 ---TPKEPA-----PTTKPSPTTPK-----EP-- 481
Db 1327 SAASLTGSKRLRTPKRGSEVPEDLAGPIELFQTPSHTKESMTNEKTKVYSRASQDPL 1385
Qy 482 ---APTTPKSAP-----
Db 1386 VDTPTSSKPPQKRSRLKADTEBEFLAFKQTPSAGKAMHTPKPAVGEBKDINTFLGTPVQ 1445
Qy 491 ---TTTKEPA-----PT-----TTK-----SAPTT 507
Db 1446 KLDQPGNLPGSNNRQLTRKEKAQALEELTGFRELQTPCTDNTADEKTTKILCKSPQS 1505
Qy 508 PKBPSPPTTKE-----PA-----
Db 1506 DPADPTTNTKQPKRSKADVEEBFLAFRLKLTSPSAGKAMHTPKAAVGEEDINTFVGT 1565
Qy 521 ---PTTPKEPAPT-----TP-----KPK 535
Db 1566 VEKDLLGNLPGSKRPQTPKAKALEDLAGFKELFQTPGHTESMTDDKITEVSCSP 1625
Qy 536 APTTPKEPA-----
Db 1626 QPDVPKTTSSKQRLKISLGVGVKEEVLPGVGLTQSGKTTQTHRETAGDGKSIKAPKE 1685
Qy 545 ---PTTPKEPA-----
Db 1686 SAKQMLDPANYGTGMRPRTPKREBAQSLDLAGFKELFQTPDHTTESTTDDKTTKIA 1745
Qy 566 EPAP---TTPKET---APTTP-----KKLTPTT----- 587
Db 1746 SPPPSMDTPTSTRRPKTLGLKRDIVBELSALKQLTQTHDKVPGDEDKGINVRETA 1805
Qy 588 PEKLAP---TTEKEPAPT---BELA-----PTTPBPPTTPPEP 623
Db 1806 KQKLDPAASVSGKRPQTPKGAQPLDLAGLKFQTPVCDKPTTHEKTTKIAKSP 1865
Qy 624 AP---TTPKAAA-----
Db 1866 QPDVPVGTPTTFKPSKRSRLKADVEEFLALRTPSVGKAMDTPKPAGGDEKDKMAPMG 1925
Qy 633 ---PNTPKPEA-----PTTPKEPAPTTPKE 654
Db 1926 TPVQKLDLGNLPGSKRPQTPKKAQALDLAGFKELFQTPGTDKPTTDEKTTKIACKS 1985
Qy 655 PAP---TTPKET---APTTPK-----
Db 1986 PQDPPVDTPASTKQPKRNLRKADVEEFLALRKTTPSAGKAMDTPKPAVDEKKNITFV 2045
Qy 670 ---GPAPTTLKEPAPTTPKXPAP-----KELAPTTTKEPTTSTSDK----- 707
Db 2046 ETPVQKLDLGNLPGSKRQ---PQTPKKAQALEDLVGFKELFQTPGHTESMTDDKITEV 2103
Qy 708 ---PAP-----TTPKGTAPT----- 719
Db 2104 SKSPQSFESFTRSRSKORLKIPLVKVDMKEEPLAVSKLRTSGTGTQTHTEPTGDSKSI 2163
Qy 720 ---TPK---BPAPT---TPKEPA----- 733
Db 2164 KAFKESPKQILDPAASVTSRRQLTRKEKABALEDLVDFKELFAPGHTESMTIDKQT 2223
Qy 734 ---PTTPKGTAPTTLKEPAPTTPKXPAPKELAP-----TTTKGPTSTSTSDKA----- 778
Db 2224 KIPCKSPPELDTATSTKRCP-KTRPRKEVKEELSVERLTQTSQSTHTHKEPASGDE 2282
Qy 779 ---PTTPKE-----TAPTT 789
Db 2283 GIKVLQKAKKXPNVEBPSPRRPRAPKKAQPLEDLAGFTELSETSGHTQESLITAKA 2342
Qy 790 PKEPAPTTPKPAPTTPPTPTTSEVSTPTTKPTTHKSPDESTPELSAETP----- 845
Db 2343 TKIPCESPLEVVDITASTKRLRLTRVQKVQKEBPSAV-KFTQTSGETTDDAKPEAGED 2401
```

```
Qy 846 ---KALENSPKFEGVPTT-----KTPAA---TKPEMTTAK 875
Db 2402 KGIKALKESAKQTPAPAAASVTGSRRRPRAPRESAQALDLAGPKDPAAGHTESMT---D 2458
Qy 876 DKTERDLRTPPE---TTTAAPKMTKETATTTKTE---SKTATTQVTTTQ--- 925
Db 2459 DKTTIPCKSSPELEDATSSKRPRTRAQKVKBELLAAGVGLTQTSGETTHTDKBPVG 2518
Qy 926 ---DTPPFKLTLLKTTTLLAPKVTITTKTITTTIMNKPEETAKPKDRA----- 970
Db 2519 EGKGTAKAF-----QPAKRNVDADVIGSRROPRAPRKPKEKAQPLEDLASQEL 2565
Qy 971 ---TNSKATTTKPKQPKTK-----APK-KPT----- 991
Db 2566 SQTGHTTELANGAADSFTSAPKQTPDSGKPLKISRVLRAKVPFVGVGVSTRDPVKSQ 2625
Qy 992 ---STKKPKTTPVRVKPKTTP----- 1009
Db 2626 SKSNTSLPLLPFKRGGKGDSVGTGKELRCMPAPEBEIVEELPASKQORVAPRARGKSEP 2685
Qy 1010 T---PRKWTMTPELNPTRIAEAMLTQTPRPNQ---TPNSKLVEVNPKSEBAGGAGE 1062
Db 2686 VVIMKRSLTRSAKRIEPAEELNSDMKTKNEEHKLQDSVPENKGISLRSRQDKTEAQ 2745
Qy 1063 TPHMLL-----REHVFMPEVTPDMDYLPRVNPQGIILNP 1096
Db 2746 ITEVFVLAERIEINREKCP---MKTSPEMD-----IQNP 2777

RESULT 13
T25697
hypothetical protein F16F9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T25697
R:Fulton, B.
submitted to the EMBL Data Library, August 1996
A:Description: The sequence of C. elegans cosmid F16F9.
A:Reference number: Z20071
A:Accession: T25697
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1229 <FUL>
A:Cross-references: UNIPROT:Q94185; EMBL:U67956; FIDN:AA07691.1; GSPDB:GNC0028; CESP:F1
A:Experimental source: strain Bristol N2; clone F16F9
C:Genetics:
A:Gene: CESP:F16F9.2
A:Map position: X
A:Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match 14.9%; Score 871.4; DB 2; Length 1229;
Best Local Similarity 25.4%; Pred. No. 2.4e-10;
Matches 346; Conservative 100; Mismatches 360; Indels 556; Gaps 66;

Qy 2 AKWTLPIYLLLSLVFVIOQVSSQELSCG-RCFESFERGRCDCDAQCKKYDKCCPDYE 60
Db 4 AW-VVSPAFILGNVQVSSLLSKTINSQSRDFKIVGHKKNCTCSC---KCVPD-- 56
Qy 61 SFCAEVHNPTSPSSKKAPPPSGASOTIKSTTRSPKPNKKTKKVIESEETEEHSVS 120
Db 57 ---AFSN---PFVSVTIISSN-----N 73
Qy 121 ENGESSSSSSSSSSTIW-----KIKSKNSAANRELQKLVKDNKNKRTK 169
Db 74 DNVDIGSGDSNPTGSS-WFOEIEATVGGQTVKSEHNIDSSVEVEKKV----- 120
Qy 170 KPFPKPVVDEAGSLDNGDFKVTTPDTSTQ---HNKYSTSPKI----- 211
Db 121 ---TISTDASTNAPTGGKSTTPEITGIVVINSKSESVT 158
Qy 212 ---TTAKPINRPSLPNSDTSKETSILVNNKETTIVETKTTTNNKQSTDGKEKT 263
Db 159 DMSITRSTTLSP-----TTELJTSPELTVSTDSSTST-EQTSPDNTTEI 202
```


Db 1406 ALKKLTQTSGETHTDVKVGGGDKSINAFRETAQKQKLDPAASVTSKRRHPT-KEKAQPL 1464
QY 412 -----PTTPK-----EPTPT-TPKEPAPTTPKE-----432
Db 1465 EDLAGWKELFQFVCTDKEFTTHEKTKIACRSQPDVDTPTSSKQSKSLRKVDVEEF 1524
QY 433 -----PART-----TPKEPA--442
Db 1525 FALRKRTPSAGKAMHTPKPAVSGEKNIYAFMGTPVQKLDLTENLTGSKRLQTTPKRAQA 1584
QY 443 -----PTAPKPA-----450
Db 1585 LEDLAGFKELFQTRGHTBESMTDKATAKACKSSQPDLDKXNPASSKRRLKTSLGKGVKE 1644
QY 451 -----PTTPKEPAPT-----TPKEPA 466
Db 1645 ELLAVGKLTQTSGETHTHTTEPTGDKSKAFWESPKQLDGAASLTGSKRLQTPKPGKS 1704
QY 467 -----PTTKESPTTPK-----BP-----APTTPKSA-490
Db 1705 EVPEDLAGFIELFQFPSHTKE-SMTNEKTKVSYRASQPDLDVDTTSSKQPKRSLRAD 1763
QY 491 -----TTTK 494
Db 1764 TBEFLAFKQTPSAGKAMHTPKPAVGEKQDINTFLGTPVQKLDQGNLPGSNRRLQTRK 1823
QY 495 EPA-----PT-----SPTTPKESPTTPKE-----518
Db 1824 EKAQALELTGTFRELFQTPCTDNPTADEKTKKILCKSPQSDPAPTNTKQRPKRSKK 1883
QY 519 -----PA-----PTT 523
Db 1884 ADVEEFLAFKLTSPSAGKAMHTPKAAVGEKQDINTFVGTPEVKQLDGLNLPGRSRPQT 1943
QY 524 PKEPAPT-----TP-----KKPAPTTPKEPA-----544
Db 1944 PREKAKALEDLAGFKELFQTPGHTBESMTDDKITEVSKSPQPDVKTPTSSKQBLKLSL 2003
QY 545 -----P 545
Db 2004 GKVGKVEVLVPGKLTQTSGETHTQHRETAGDGKSIKAFKESAKQWLDPAVNYGTGWERWP 2063
QY 546 TTPKEPA-----PTTKKAPAPTAPKEPAP-----TTPKET--APT 579
Db 2064 RTPKEAQSLEDLAGFKELFQTPDHTTESTTDDTKTKIACKSPFPESMDTPTSTRRPKT 2123
QY 580 P-----KKLTPTT-----PEKLAP-----TTPKEPAPT 603
Db 2124 PLGKRDIVEELSALKQLTQTHTDKVPGEDGKINVFRETAKQKLDPAASVTSKRRQPT 2183
QY 604 P-----BELA-----PTTPBEPPTTPBEPAPT-----626
Db 2184 PKGKAQPLEDLAGFKELFQTPVCTDKPTTHEKTKIACRSQPDVGTPTIFPKQSKRSL 2243
QY 627 -----TPKAAA-----P 633
Db 2244 RKADVEESLALRKRTPSVGKAMDTPKPAAGDEKQKMAFMGTFVQKLDLPGNLPGSKRWP 2303
QY 634 NTPKEPA-----PTTPKEPAPTTPKEPAP-----TTPKET-----663
Db 2304 QTFKEKAQALEDLAGFKELFQTPGTDKPTTDEKTKIACKSPQPDVDTPTASTKQRPXRN 2363
QY 664 -----APTPPK-----GTAPTTLKE 678
Db 2364 LRKADVEEFLALRKRTPSAGKAMDTPKPAVSDKRNINTFVETPVQKLDLGNLPGSKRQ 2423
QY 679 PAPTTPKPKAP-----KELAPTTKETPTSTSDK-----PAP-----710
Db 2424 --POTPKKAEALEDLVGFKELFQTPGHTBESMTDDKITEVCKSPQESFKTSRSSQR 2481
QY 711 -----TTPKGTAPT-----TPK--EPAPT--727

Db 2482 LKIPLVKVDKBERPLAVSKLRTSGGETTQHTTEPTGDSIKAFKSPKQILDPAASVTG 2541
QY 728 -----TPKEPA-----PTTPKGTAAPTTL 745
Db 2542 SRRQLRTRKEKARALEDLVDFKELFSAPGHTBESMTIDKNTKIPCKSPPELTDATSTK 2601
QY 746 KEPAPTTPKPKAPKELAP-----TTTKGPTSTTSKPA-----778
Db 2602 RCP-KTRPRKEVKEELSVAVERLTQTSQSTHKEPASGDEGKVLKQRAKKPNPVEEE 2660
QY 779 -----PTTPKE-----TAPTPKEPAPTTPKPAPTTPET 808
Db 2661 PSRRPRAPKEKAQPLEDLAGFTSELSETSGHTQESLTAGKATKIPCESPPLVVDITAST 2720
QY 809 PPPTTSVSPTTKETTKTHKSPDESTPELSAEP-TP-----KALENSPKPGVPTT-860
Db 2721 KRHLRTRVQVQVKEEPSAV-KFTQTSGETTDAKPEGEDKGKALKESAKOTPPAPAS 2779
QY 861 -----KTPAA--TKPEMTTAKDKTTERDLRTPE--TTT 891
Db 2780 VTGSRPRAPRESAQAIEDLAGFKDPAAGHTEESMT--DDKTKIPCKSSPELEDTAT 2836
QY 892 AAPKMTKETATTEKTTT-----SKITATTTQVSTTTQ-----DTTPFKITILKTTTLA 941
Db 2837 SSKERPRTRAQKVEKSELLAVGKLTQTSGETTHTDKEPVGEGKGTAKFK-----2886
QY 942 PKVTTTKTTTITIMNKPBEETAKPKDRA-----T 971
Db 2887 --QPAKRVDAEDVIGSRQRPAPKEKAQPLEDLASQELSQTPGHTTELANGAADSFT 2943
QY 972 NSKATTPKQPKTK-----APK-KPT-----991
Db 2944 SAPKQTPDSGKPLKISRRLRAPKVEPVGVWSTDPVKSQSKSNTSLPLPFKRGCGKD 3003
QY 992 -----STKKPKTMRVRKPTTPT--PRKMTSTWPELNPTS 1025
Db 3004 GSVTGTGRLKCMAPAPESIVEELFASKQVAPAPRAGKSEPVVMKSLTSKRIEPAE 3063
QY 1026 RIABAMLQTTTRPNQ-----TPNSKLVEVNPKESDAGAGETPHMLL-----R 1069
Db 3064 ELNSNDMKTKEEHKLDQSVPENKGISLRSRQDKTEAEQOITEVFLAERIEINRNEKK 3123
QY 1070 PHVEMPEVTDMYDLPRVFNQGIINP 1096
Db 3124 P-----MKTSPEMD-----IQNP 3136
RESULT 15
Ti18535
high molecular mass nuclear antigen - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18535
R:Shimada, K.; Hazata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A>Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick
A:Reference number: Z18955; MUID:9803440; PMID:9365273
A:Molecule type: mRNA
A:Residues: 1-1151 <SH1>
A:Cross-references: UNIPROT:O57580; EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BAA241

Query Match 14.8%; Score 860.7; DB 2; Length 1151;
Best Local Similarity 22.2%; Pred. No. 3.5e-10;
Matches 298; Conservative 89; Mismatches 360; Indels 593; Gaps 58;
QY 69 PTSPSSK-----KAPPPSG-----ASQTI 88
Db 27 PTAPRKPWPIAELHPAAQPPKWPVIGAPPPCTEPTTPSKPTDGADAAPKASAE 86
QY 89 KSTTKRSPKPKKTKKVKIESEITEHSVSNQESSSSSSSSSSSTTWIKSKNSA 148

```
Db 87 TSPPPASPPDPGPK-----APSGAGEAEGT----- 113
Qy 149 ANRELQKLVKXDNKKNRTKKPTK-----PPVDEAG-----SGLDNG-DF 190
Db 114 -----PPSQGAPGPPPSOGAAGPKGDGTAPQSPGKSGADG 151
Qy 191 KVTTPDTSTTOHNVKSTGPKITTAKEPINSRPLPNSDTSKETSLTVNKETTVETKETT 250
Db 152 KPAQDV-----PKAITA-ATEAR-----ASAAPTVPKATA----- 183
Qy 251 TNKQSTSDGKXTTSKETSQSTIEK-----TSKDLAPTSKVLAKP----- 290
Db 184 -----EATAVTAASQSAKPAATDAAAVTAASQSAKPAATVEVKFAAAVAKEAKAV 233
Qy 291 -----TPKAETTKGPAULTTPKEP-----TPTTKEPASTTPKEPTPTTIKSA 333
Db 234 TAAAAAPKATAEAKPAPVTSPTIPCSSAEAKPLTAASPTASKATAEAKPVPATASLAKTK 293
Qy 334 PTPKEPAPTTTKSAPTTTPKEPAPTTK-----EPAPTTPE 370
Db 294 VTAEAKPAP-----SPSVPK-ATTDKAVTATAPKAGPDVKPAAVCAEAKPAPPPPPQ 346
Qy 371 PAPTTTKEPAPTTTKSAPTT-----PKEPA-----PTTPKAPAPTTKEPAPTTPE 417
Db 347 QLPKAAAAAAPTGTLEKATAPPHGSPRANGHTVTPNVPRAAAATVP--TAGAVPKA 404
Qy 418 PTPPTTKEPAPTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKEPSPTT 477
Db 405 STGTP-----AAAPQOPV-----KAAVTPPSPOQAVPR-----AATAAAAPVT 445
Qy 478 PKEPAPTTTKSAPTTKEPAP-----TTKSAPTTPKEPSPT--ITKEPAPTTKEP- 527
Db 446 POOP---VTKAATTTNATPPPOPIPKAATTTTATVTPQPIPKAGIDAAAPPAPVPKAPS 502
Qy 528 -----APTPKAPAPTTKEPAPTTTPKEPAPTTTKKPA-----PTAP-----KE 566
Db 503 DGRAATPGVNAATDPQKPPPTQSPVSAVTEPKPEQRAAPPSPNEATPAVPSPSNLKS 562
Qy 567 PAPTPPK-----ETAPTPPKLUTPTTPEKLAFTPEKAPAPTTPELA 608
Db 563 PLFTIPKVPFLMALTFQVTAQMTQLAATKPSIVPKASPK-ALMTPPPPPPGLPRALA 621
Qy 609 -----PTTP-----EPTPTTP-----EAPPTTPKAA----- 631
Db 622 AAKLGLSPSPVSAAMHAKVTRPLPASPVPMAASPSASLGPDAARVALATNAASPGAKPE 681
Qy 632 ----- 631
Db 682 AAGNGTILMAPGAANTOMAPIGAAGAAQATAPMGAAATHVSPMGAGGATOMSPGTGAANTH 741
Qy 632 -----APNTPEP-----APTTPEP-----APTTPEP----- 655
Db 742 NSPIGAGGATOMSPGAANTOMSPMGATTONKSPMGAAATQPSMGAAATQVATATSAGN 801
Qy 656 -----APTTPEKAPAPTTTPKGT-----APTILKEP--APTTPKAPKELAPTTTKE 699
Db 802 TMQVSPMGAAATPPQTPSVGAATTPQPSMGAAATILMSMGAAATTPQ---PSPMGAVTTQP 858
Qy 700 P--TSSTSDKAP-----TPKGTAPTTPEP-APTTPKEPAPTTTPKGTAPTTILKEPA 749
Db 859 PMAAINTQPPPMMAASTPQSTPMGAATTPQSPMGATTONKSPMGATTONKSPMGATTONKSP 905
Qy 750 PTPPKAPKELAPTTTTPKGTSTSDKAPAPTTTPKGTAPTTT-----KEPAPTTTPKAPAPTT 805
Db 906 -STQAP-----PTVAGSPt-----PPPIPPSPPTACTSPQMSKSPPPDPKAPSAAA 953
Qy 806 PETPPPTTSEVSTPTTTPKEPTTI---HKSPDESTPELSAETPKALENSKPEPGVPTTKT 862
Db 954 QTSAPAAHVANASPGVTVASVPAPIGVTEASPSADGARLS--PGPTAATDGPK-----AS 1004
Qy 863 PAATKPEMTTAKDKTTERDLRTTPTTTPAAKMTKETTATTTTETSKITATTTQVTTST 922
Db 1005 PAAT-ADVTEAATDVTA---AATAVPAEAAPTKAKSSSSSSSSSSSSSSSSSSSSSSSSSS 1059
```

```
Qy 923 TTODTTPFKITTLKTTTLAPKVTITTKTITTEIMNKPBEATAKPKDRATNSKATTPKPK 982
Db 1060 SDSA-----SSSSSSSNPAS 1075
Qy 983 PTKAPKKPTSTKKPKTMTPRVRKPTTTPRKMSTMPELNPTSRIAEAMLOTTTRPNQTP 1042
Db 1076 PA-----PAVGDQOQOQMTPCAASQSVPP-----VTEAAVQ----- 1104
Qy 1043 NSKLVEVNPKSEDAAGAEGE 1062
Db 1105 -----EAAAAAAAAGAERE 1119
```

Search completed: October 13, 2004, 11:56:32
Job time : 64.3855 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 87.8789 Seconds
(without alignments)
4449.477 Million cell updates/sec

Title: SEQ1-B
Perfect score: 5826
Sequence: 1 MAWKTLPYLLLSVFIQ.....DMDYLPVPGIINPMLS 1090

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:.*
1: Geneseq1980s:.*
2: Geneseq1990s:.*
3: Geneseq2000s:.*
4: Geneseq2001s:.*
5: Geneseq2002s:.*
6: Geneseq2003as:.*
7: Geneseq2003bs:.*
8: Geneseq2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	5811	99.7	1299	4 AAM24322	Human EST
2	5811	99.7	1404	2 AAB26049	MSF precu
3	5811	99.7	1404	4 AAB29773	Human meg
4	5811	99.7	1404	4 AAB60568	Human meg
5	5811	99.7	1404	8 ADM98014	Human meg
6	5811	99.7	1415	4 AAU32262	Novel hum
7	5788	99.3	1404	7 ADK65839	Angiogene
8	5608.7	96.3	1311	8 ADK67912	Human ext
9	5370.2	92.2	1320	7 ADK65819	Angiogene
10	5344.6	91.7	1270	8 ADK67911	Human ext
11	2920	50.1	546	4 ABUS3252	Human tes
12	2888	49.6	551	4 ABUS3253	Human tes
13	2880.1	49.4	902	4 AAB29778	Human tes
14	2757	47.3	513	4 ABUS3254	Human tes
15	1994.2	34.2	452	2 AAB80041	Human meg
16	1409.9	24.2	538	5 AAC18834	3' cartil
17	1394.5	23.9	5179	4 AAM24516	C899P pre
18	1394.5	23.9	5179	6 ABF55365	Human col
19	1394.5	23.9	5179	6 AB007258	Human p53
20	1394.5	23.9	5179	7 ADD48091	Human Pro
21	1394.5	23.9	5179	7 ADD48091	Human Pro
22	1335.4	22.9	292	5 AAU11261	Human HAP
23	1127.2	19.3	1664	2 AAW43106	C. thermo
24	1103.7	18.9	1795	4 ABB69806	Drosophil
25	1099.6	18.9	717	4 ABUS3144	Human tes

ALIGNMENTS

RESULT 1
AAM24322
ID AAM24322 standard; protein; 1299 AA.
XX
AC AAM24322;
AC
DT 12-OCT-2001 (first entry)
XX
DE Human EST encoded protein SEQ ID NO: 1847.
XX
KW Human; sheep; pig; cow; fruit fly; Yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;
KW gene therapy; nutrition.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US002687.
XX
PR 25-JAN-2000; 2000US-00491404.
PR 17-JUL-2000; 2000US-00617746.
PR 03-AUG-2000; 2000US-00631451.
PR 15-SEP-2000; 2000US-00663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI; 2001-476164/51.
DR N-PSDB; AAH98981.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use.
XX
PT Claim 20; Page 1198-1201; 1275pp; English.
XX
PS The present invention provides the protein and coding sequences of novel
PS proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a

26	1097	18.8	214	4	ABUS3255	Human tes
27	1091.9	18.7	8991	6	ABU08487	S. pneumo
28	1053.7	18.1	763	3	NAG38942	Arabidops
29	1039.7	17.8	1049	4	ABB61364	Drosophil
30	1031	17.7	2284	4	ABB71434	Drosophil
31	1027	17.6	5703	8	ADL23265	Human MUC
32	1010.7	17.3	188	5	AAO18833	5' cartil
33	993.1	17.0	2112	4	ABB60403	Drosophil
34	987.9	17.0	770	4	ABUS3141	Human tes
35	987.1	16.9	4315	5	ABP43908	MUC5B par
36	981	16.8	778	4	ABUS3143	Human tes
37	974.4	16.7	692	4	ABUS3155	Human tes
38	964.1	16.5	717	4	ABUS3150	Human tes
39	964.1	16.5	717	4	ABUS3149	Human tes
40	964.1	16.5	717	4	ABUS3151	Human tes
41	964.1	16.5	717	4	ABUS3145	Human tes
42	964.1	16.5	717	4	ABUS3148	Human tes
43	964.1	16.5	717	4	ABUS3147	Human tes
44	964.1	16.5	717	4	ABUS3146	Human tes
45	960.8	16.5	1460	7	ADN39110	Cancer/an

XX Claim 1, 2 and 3; Fig 1; 87pp; English.

XX The sequence given is a full length translation from the megakaryocyte

PS stimulating factor (MSF) precursor. The sequence covered by exons II, III

CC and IV encodes megakaryocyte stimulating factor (MSF). This sequence is

CC modified by the addition of an N-terminal sequence encoding a secretory

CC leader, an initiating methionine preceding exon II and a terminating

CC codon following exon IV. The cDNA sequence given contains sequences

CC derived from human megakaryocyte colony stimulating factor (meg-CSF).

CC Exon I contains the initiating methionine, and encodes a classical

CC mammalian protein secretion signal sequence. The sequence encoding the

CC original meg-CSF includes exons II-IV and is thought to terminate in the

CC region between amino acid residues 134 - 147. The primary transcript of

CC this gene may be cleaved in different ways to yield a family of mRNA's

CC each encoding a different MSF protein. Exons V and VI are thought to be

CC related to the activity of the factor and are also implicated in the

CC stability, folding and processing of the molecule. These exons are also

CC thought to play a role in the observed synergy of MSF with other

CC cytokines. Exons V - XII are believed to be implicated in the processing

CC or folding of the appropriate structure of the resulting factor, i.e. one

CC or more of these exons may contain sequences which direct proteolytic

CC cleavage, adhesion, organisation of the cellular matrix or extracellular

CC matrix processing. Both naturally occurring and non-naturally occurring

CC MSF's may be characterised by various combinations of alternatively

CC spliced exons from this sequence, with the exons spliced together in

CC differing orders to form different members of the MSF family. (Updated on

CC 25-MAR-2003 to correct FN field.)

XX

SQ Sequence 1404 AA;

Query Match 99.7%; Score 5811; DB 2; Length 1404;

Best Local Similarity 95.6%; Pred No. 6.8e-159;

Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGCGEGYSDATCNDCYNQCHYMECCPDF 60

DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGCGEGYSDATCNDCYNQCHYMECCPDF 60

QY 61 KRVCTAELSCKRCFESFERGECDDAOCKYDCCPDYESFCA----- 105

DB 61 KRVCTAELSCKRCFESFERGECDDAOCKYDCCPDYESFCAEVNHNPPSPSKKAP 120

QY 106 -----EHSVSVNQSSSSSSSSSSSSSSSIW 130

DB 121 PPSGASQTIKSTKRSPPKPKTKTKVIESEITEHSVSVNQSSSSSSSSSSSIW 180

QY 131 KIKSSKNSAANRELQKLVKONKQRTKKPTPPPPVVDVDEAGSLDNGDFKVTPTDST 190

DB 181 KIKSSKNSAANRELQKLVKONKQRTKKPTPPPPVVDVDEAGSLDNGDFKVTPTDST 240

QY 191 TOHNKVSTSPKTTAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTINKGTSIDG 250

DB 241 TOHNKVSTSPKTTAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTINKGTSIDG 300

QY 251 KEKTTSAKTSQIEKTSKADLAPTSKVLAKPTPKABTTKGPALTTKPEPTTTPKEPAS 310

DB 301 KEKTTSAKTSQIEKTSKADLAPTSKVLAKPTPKABTTKGPALTTKPEPTTTPKEPAS 360

QY 311 TTPKEPTPTTIKSAPTTTKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 370

DB 361 TTPKEPTPTTIKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 420

QY 371 APTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTK 430

DB 421 APTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTK 480

QY 431 EPAPTAPKPAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 490

DB 481 EPAPTAPKPAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 540

QY 491 TTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 550

Db 541 TTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 600

QY 551 APTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 610

Db 601 APTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 660

QY 611 PEPAPTTAKAAPTNPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 670

Db 661 PEPAPTTAKAAPTNPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 720

QY 671 APTTPKPAKELAPTTTKEPPTTSTSDKPAPTTTPKGTAPTTTPKEPAPTTTKEPAPTTTPKG 730

Db 721 APTTPKPAKELAPTTTKEPPTTSTSDKPAPTTTPKGTAPTTTPKEPAPTTTKEPAPTTTPKG 780

QY 731 TAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 790

Db 781 TAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 840

QY 791 KPAPTTPETPPPTTSEVSTPTTKEPTTIHKSPDESTPELSAEPPTPKALENSPKPEGVPT 850

Db 841 KPAPTTPETPPPTTSEVSTPTTKEPTTIHKSPDESTPELSAEPPTPKALENSPKPEGVPT 900

QY 851 TKTPAATKPEMTTAKDKTTERDLRTPPTTAAAPKXKETAATTTESKITATTTQV 910

Db 901 TKTPAATKPEMTTAKDKTTERDLRTPPTTAAAPKXKETAATTTESKITATTTQV 960

QY 911 TSTTTQDTPFKITTLTKTTLAKPVTTTKTITTTTEIMNKPEETAKEKDRATNSKATTPK 970

Db 961 TSTTTQDTPFKITTLTKTTLAKPVTTTKTITTTTEIMNKPEETAKEKDRATNSKATTPK 1020

QY 971 PQKETKAPKPTSTKPKTMFVRKPKTTPTRKMTSTMPELNPTSSIAEAMLQTTTRPN 1030

Db 1021 PQKETKAPKPTSTKPKTMFVRKPKTTPTRKMTSTMPELNPTSSIAEAMLQTTTRPN 1080

QY 1031 QTPNSKLVEVNPXSDEAGGAGETPHMLRPHVFMPEVTPDMOYLPRVNOGIIINPMLS 1090

Db 1081 QTPNSKLVEVNPXSDEAGGAGETPHMLRPHVFMPEVTPDMOYLPRVNOGIIINPMLS 1140

RESULT 3

AAB29773

ID AAB29773 standard; protein; 1404 AA.

XX

AC AAB29773;

XX

DT 28-FEB-2001 (first entry)

XX

DE Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.

XX

KW Human MSF; megakaryocyte stimulating factor; tribonectin;

KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;

KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;

XX friction coefficient reduction; gene therapy; antiarthritic; osteopathic.

OS Homo sapiens.

XX

PN WO200064930-A2.

XX

PD 02-NOV-2000.

XX

PZ 24-APR-2000; 2000WO-US010953.

XX

PR 23-APR-1999; 99US-00298970.

XX

PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

XX

PI Jay GD;

XX

XX WPI; 2001-024673/03.

DR N-PSDB; AAC81498.

XX

PT Novel tribonectin polypeptide useful as lubricant for treating

PT osteoarthritis, comprises O-linked lubricating moiety.

XX PS Claim 3; Page 7; 47pp; English.

CC The invention relates to a human tribonectin which is a product of

CC alternative splicing of the human MSF (megakaryocyte stimulating factor)

CC gene. The tribonectin has at least one O-linked oligosaccharide

CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats

CC of a motif having at least 50% identity to the sequence KEPAPTT

CC (AAB29774). The invention also relates to a nucleic acid encoding a human

CC MSF-derived tribonectin; a biocompatible composition comprising a human

CC tribonectin for inhibiting tissue adhesion formation; and a method of

CC diagnosing osteoarthritis or a predisposition to osteoarthritis by

CC measuring the amount of MSF or its fragment in a biological sample of a

CC mammal, wherein an increased amount of MSF compared to a control

CC indicates the presence of or predisposition to developing osteoarthritis.

CC The tribonectin and DNA encoding it are useful in the treatment of

CC osteoarthritis, where they may be used for lubricating mammalian joints,

CC such as articulating joints of humans, dogs or horses. The tribonectin,

CC when formulated as a membrane, foam, gel or fibre, is useful for

CC inhibiting adhesion between two surfaces such as the injured tissues of a

CC mammal, where the injury is caused by a surgical insertion or trauma, or

CC an artificial device e.g., an orthopaedic implant. In particular, one of

CC the surfaces is pericardial tissue. DNA encoding a tribonectin may be

CC used in gene therapy. The present sequence represents human MSF

XX XX

XX Query Match 99.7%; Score 5811; DB 4; Length 1404;

XX Best Local Similarity 95.6%; Pred. No. 6.8e-159;

XX Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKLTPIYLLLLLVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYNECCPDF 60

DB 1 MAWKLTPIYLLLLLVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYNECCPDF 60

QY 61 KRVCTAELSCKRCFESFERGECDCDAQCKYDKCCPDYBSFCA----- 105

DB 61 KRVCTAELSCKRCFESFERGECDCDAQCKYDKCCPDYBSFCAEVHNPTSPSSKKAP 120

QY 106 -----EHSVSENQESSSSSSSSSSSSSTIIV 130

DB 121 PPSGASQTIKSTTKRSPKPPNKKTKVIESEITEHSVSENQESSSSSSSSSTIIV 180

QY 131 KTKSSKNAANRELQKLVKDNKNRTKKPTKPPVVDAGSLDNGDFKVTTPDTST 190

DB 181 KTKSSKNAANRELQKLVKDNKNRTKKPTKPPVVDAGSLDNGDFKVTTPDTST 240

QY 191 TOHNKVTSPKITTAKPINRPSLPNNSDTSKETSLTVNKETTVETKTTTNKQSTDG 250

DB 241 TOHNKVTSPKITTAKPINRPSLPNNSDTSKETSLTVNKETTVETKTTTNKQSTDG 300

QY 251 KEKTTSAKETOSIETSKADLAPTSKVLAKEPTKPAETTTKGPALTTPKEPTTPKEPAS 310

DB 301 KEKTTSAKETOSIETSKADLAPTSKVLAKEPTKPAETTTKGPALTTPKEPTTPKEPAS 360

QY 311 TTPKEPTTTIKSAPTTKPEPAPTTTKSAPTTTPKEPAPTTTKPEPAPTTTKPEP 370

DB 361 TTPKEPTTTIKSAPTTKPEPAPTTTKSAPTTTPKEPAPTTTKPEPAPTTTKPEP 420

QY 371 APTTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPTTTTPKEPAPTTTKPK 430

DB 421 APTTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPTTTTPKEPAPTTTKPK 480

QY 431 EPAPTAPKKPAPTTKPEPAPTTKPEPAPTTTKPSPTTPKEPAPTTTKSAPTTTKPEAPT 490

DB 481 EPAPTAPKKPAPTTKPEPAPTTKPEPAPTTTKPSPTTPKEPAPTTTKSAPTTTKPEAPT 540

QY 491 TTKSAPTTKPEPSPTTTKPEPAPTTKPEPAPTTKPKPAPTTTPKEPAPTTTKPK 550

DB 541 TTKSAPTTKPEPSPTTTKPEPAPTTKPEPAPTTKPKPAPTTTPKEPAPTTTKPK 600

QY 551 APTAPKEPAPTTKPEAPTTPKKTTPPTPEKLAPTTPEKAPTTPEELAPTTPEEPTPTT 610

DB 601 APTAPKEPAPTTKPEAPTTPKKTTPPTPEKLAPTTPEKAPTTPEELAPTTPEEPTPTT 660

QY 611 PEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTTLKEP 670

DB 661 PEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTTLKEP 720

QY 671 APTTPKPKAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTTPKEPAPTTTPKEPAPTTPKG 730

DB 721 APTTPKPKAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTTPKEPAPTTTPKEPAPTTPKG 780

QY 731 TAPTTTKEPAPTTPKAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTTPKEPAPTTTPKEPAPTTPK 790

DB 781 TAPTTTKEPAPTTPKAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTTPKEPAPTTTPKEPAPTTPK 840

QY 791 KPAPTTTPETPPPTTSVSTPTTKEPTTIHKSPDSTPELSAETPKALENSPKPGVPT 850

DB 841 KPAPTTTPETPPPTTSVSTPTTKEPTTIHKSPDSTPELSAETPKALENSPKPGVPT 900

QY 851 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAAPMKETATTTTEKTESKITATTQV 910

DB 901 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAAPMKETATTTTEKTESKITATTQV 960

QY 911 TSTTTQDTPPFKITTLLKTTTLAPKVTITTKTITTTIMNKPEETAKPKDRATNSKATTPK 970

DB 961 TSTTTQDTPPFKITTLLKTTTLAPKVTITTKTITTTIMNKPEETAKPKDRATNSKATTPK 1020

QY 971 PQKPTKAPKPTSTKKPKTMPRVKPTTPTRKMTSTMPELNPTSRIAEAMLQTTTRPN 1030

DB 1021 PQKPTKAPKPTSTKKPKTMPRVKPTTPTRKMTSTMPELNPTSRIAEAMLQTTTRPN 1080

QY 1031 QTPNSKLVEVNPKSDAGAGETPHMLLRPHVFEVPTPDMDYLPVPNOGIIINPMLS 1090

DB 1081 QTPNSKLVEVNPKSDAGAGETPHMLLRPHVFEVPTPDMDYLPVPNOGIIINPMLS 1140

RESULT 4

AAB60568

ID AAB60568 standard; protein; 1404 AA.

XX AC AAB60568;

XX DT 27-APR-2001 (first entry)

XX XX Human megakaryocyte stimulating factor (MSF, CACP).

DE XX Human; CACP protein; camptodactylly-arthropathy-coxa vara-pericarditis;

XX KW MSF; megakaryocyte stimulating factor; synovial lubricant;

KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;

XX KW antiarthritic.

OS Homo sapiens.

XX WO200107068-A1.

PN XX

PD 01-FEB-2001.

XX XX 21-JUL-2000; 2000WO-US020002.

XX PF 23-JUL-1999; 99US-0145328P.

PR 19-JUL-2000; 2000US-00145328.

XX XX (UYCA-) UNIV CASE WESTERN RESERVE.

XX PA Warman ML;

PI WPI; 2001-182721/18.

XX XX New composition comprising the camptodactylly-arthropathy-coxa vara-

PT pericarditis protein in combination with an anesthetic, useful for

PT treating osteoarthritis, or as lubricants of tissue and joints.

XX XX Example 1; Page; 34pp; English.

XX PS

CC The invention relates to a method of treating osteoarthritis via the
CC administration of a composition comprising the camptodactyl-arthropathy-
CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
CC The composition may further comprise a local anaesthetic. The composition
CC of the invention may be administered via intra-articular or intravenous
CC injection. The human CACP protein is identified in the invention as being
CC megakaryocyte stimulating factor (MSF). The gene encoding CACP protein
CC (MSF) is located on chromosome 1q25-31, and mutations in this gene are
CC responsible for the heritable disorder camptodactyl-arthropathy-coxa
CC vara-pericarditis, in which patients have synovial hyperplasia without
CC evidence of inflammation. CACP protein (MSF) acts as a synovium
CC lubricant, and can be used to lubricate tissue and joints in the
CC treatment of osteoarthritis. The composition may be applied to reduce the
CC symptoms of osteoarthritis (e.g., joint pain, loss of range of movement
CC or joint damage). The present sequence represents human megakaryocyte
CC stimulating factor (MSF, CACP protein). Note: this sequence is not given
CC in its entirety in figure 4 of the specification, although a GenBank
CC accession number was given. This sequence was therefore obtained from
CC GenBank (U70316)
XX
SQ Sequence 1404 AA;

Query Match 99.7%; Score 5811; DB 4; Length 1404;
Best Local Similarity 95.6%; Pred. No. 6.8e-159;
Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFIQVSSODLSSCAGCGEGYSRATCNCYNCOHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFIQVSSODLSSCAGCGEGYSRATCNCYNCOHYMECCPDF 60
QY 61 KRVCTAELSCGRCPESFERGECDCDAQCKYKCCPDYSEFCA----- 105
DB 61 KRVCTAELSCGRCPESFERGECDCDAQCKYKCCPDYSEFCAEVHNPTSPSPSKKAP 120
QY 106 -----EHSVSENOESSSSSSSSSSSSSTI 130
DB 121 PPSGASQIKSTTKSPKPPNKKTKVISEITEHSVSENOESSSSSSSSSSSTI 180
QY 131 KIKSKNSAANRELQKLVKDNKNKRTKKPTKPPVVDKAGSLDNGDFKVTPTDST 190
DB 181 KIKSKNSAANRELQKLVKDNKNKRTKKPTKPPVVDKAGSLDNGDFKVTPTDST 240
QY 191 TQHNKVSPTKITTAKPINRPSLPPNSDTSKETSITVKNKETTVEKTTTNNKQTSIDG 250
DB 241 TQHNKVSPTKITTAKPINRPSLPPNSDTSKETSITVKNKETTVEKTTTNNKQTSIDG 300
QY 251 KEKTSKETSQIEKTSADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPKBPAS 310
DB 301 KEKTSKETSQIEKTSADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPKBPAS 360
QY 311 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKEP 370
DB 361 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKEP 420
QY 371 APTTKSAPTTKPEAPTTKPKAPTTKPEAPTTKPKPTPTTTPKPEAPTTKPEAPTTTPK 430
DB 421 APTTKSAPTTKPEAPTTKPKAPTTKPEAPTTKPKPTPTTTPKPEAPTTKPEAPTTTPK 480
QY 431 EPAPTAPKPKAPTTKPEAPTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPT 490
DB 481 EPAPTAPKPKAPTTKPEAPTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPT 540
QY 491 TTKSAPTTKPEAPTTTKPEAPTTKPEAPTTKPKAPTTKPEAPTTKPEAPTTTKPK 550
DB 541 TTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTKPKAPTTKPEAPTTKPEAPTTTKPK 600
QY 551 APTAPKPEAPTTKPEAPTTTKPKLPTTTPTEKLAPTTTPEKAPTTTPEELAPTTPEPTPTT 610
DB 601 APTAPKPEAPTTKPEAPTTTKPKLPTTTPTEKLAPTTTPEKAPTTTPEELAPTTPEPTPTT 660
QY 611 PPEAPTTKAAAPNTKPEAPTTKPEAPTTKPKAPTTKPKAPTTKPEAPTTKGTAPTTTKEP 670
DB 661 PPEAPTTKAAAPNTKPEAPTTKPEAPTTKPKAPTTKPKAPTTKPEAPTTKGTAPTTTKEP 720

QY 671 APTTKPKKAPKELAPTTTKEPTSTTSKDPAPTPKGTAPTTKPEAPTTTKEPAPTTPKG 730
DB 721 APTTKPKKAPKELAPTTTKEPTSTTSKDPAPTPKGTAPTTKPEAPTTTKEPAPTTPKG 780
QY 731 TAPTTLKEPAPTTPKPKAPKELAPTTTKEPTSTTSKDPAPTPKGTAPTTTKEPAPTTPK 790
DB 781 TAPTTLKEPAPTTPKPKAPKELAPTTTKEPTSTTSKDPAPTPKGTAPTTTKEPAPTTPK 840
QY 791 KPAPTTPPTPTTSEVSTPTTKEPTTHKSPDESTPDSABPTPKALENSPKERGVT 850
DB 841 KPAPTTPPTPTTSEVSTPTTKEPTTHKSPDESTPDSABPTPKALENSPKERGVT 900
QY 851 TKTPAATKPMETTTAKDKTTERDLRTTPTTAAAPKMTKETATTTTKEKTTESKITATTQV 910
DB 901 TKTPAATKPMETTTAKDKTTERDLRTTPTTAAAPKMTKETATTTTKEKTTESKITATTQV 960
QY 911 TSTTTQDTPFKITTLTKTTTLAPKVTTKKTIITTTIMNKPEETAKPKORATNSKATTPK 970
DB 961 TSTTTQDTPFKITTLTKTTTLAPKVTTKKTIITTTIMNKPEETAKPKORATNSKATTPK 1020
QY 971 POKPTKAPKPTSTKPKTMPVRKPTTPTPEKMTSTMPELNPTSRIAEAMLOTTTRPN 1030
DB 1021 POKPTKAPKPTSTKPKTMPVRKPTTPTPEKMTSTMPELNPTSRIAEAMLOTTTRPN 1080
QY 1031 QTENSKLVEVNPKSEDAGGAEGETPHMLLRPHVMEPTVDMDYLRVFNQGIINPMLS 1090
DB 1081 QTENSKLVEVNPKSEDAGGAEGETPHMLLRPHVMEPTVDMDYLRVFNQGIINPMLS 1140
RESULT 5
ADM98014
ID ADM98014 standard; protein; 1404 AA.
XX
AC ADM98014;
XX
DT 01-JUL-2004 (first entry)
XX
DE Human megakaryocyte stimulating factor (MSF).
XX
KW lubricating polypeptide; O-linked oligosaccharide; joint lubrication;
KW CAP; camptodactyl-arthropathy pericarditis; osteoarthritis; human;
KW megakaryocyte stimulating factor; MSF.
XX
OS Homo sapiens.
XX
FN US2004072741-A1.
XX
ED 15-APR-2004.
XX
PF 02-JUL-2001; 2001US-00897188.
XX
PR 23-APR-1999; 99US-00298970.
PR 24-APR-2000; 2000US-00556246.
XX
PA (JAYG/) JAY G D.
XX
PI Jay GD;
XX
DR WPI; 2004-373948/35.
DR N-PSDB; ADM98015.
XX
PT New tribonectin polypeptides and polynucleotides for lubricating joints
PT or other tissues to prevent or treat Camptodactyl-arthropathy-
PT pericarditis syndrome or osteoarthritis.
XX
PS Claim 1; SEQ ID NO 1; 34pp; English.
XX
CC The invention relates to a lubricating polypeptide and at least one O-
CC linked oligosaccharide. The composition and methods are useful for
CC lubricating joints or other tissues to prevent or treat camptodactyl-
CC arthropathy pericarditis (CAP) or osteoarthritis in mammals. The present
CC sequence represents the amino acid sequence of the human megakaryocyte

Db	12	MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNOCHYMECCPDF	71
Qy	61	KRVCTAELSCKGRCFESFERGECDCDAQCKKYDKCCPDYESFCA-----	105
Db	72	KRVCTAELSCKGRCFESFERGECDCDAQCKKYDKCCPDYESFCAEVNPTSPSSKKAP	131
Qy	106	-----EHSVSENQESSSSSSSSSSSTIW	130
Db	132	PPSGASQTIKSTTKRSPKPNKKTKVIESBEITEHVSSENQESSSSSSSSSSSTIW	191
Qy	131	KTKSSKNSAANRELQKKLVKDNKNRKKKTPKPPVVDAGSLDNGDFKVTTPDST	190
Db	192	KTKSSKNSAANRELQKKLVKDNKNRKKKTPKPPVVDAGSLDNGDFKVTTPDST	251
Qy	191	TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSILTVNKETTVTKETTTNKQTSIDG	250
Db	252	TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSILTVNKETTVTKETTTNKQTSIDG	311
Qy	251	KEKTTSAKTSQSIKTSKADLAPTSKVLAKPTPKAETTTKGPAITTPKEPTTPKEPAS	310
Db	312	KEKTTSAKTSQSIKTSKADLAPTSKVLAKPTPKAETTTKGPAITTPKEPTTPKEPAS	371
Qy	311	TTPKETPTTIKSAPTTKEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTKPEPAPTTTKEP	370
Db	372	TTPKETPTTIKSAPTTKEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKEPAPTTTKEP	431
Qy	371	APTTKSAPTTKEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKEPAPTTTKEPAPTTK	430
Db	432	APTTKSAPTTKEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKEPAPTTTKEPAPTTK	491
Qy	431	EPAPTAKKAPATTKEPAPTTKPEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT	490
Db	492	EPAPTAKKAPATTKEPAPTTKPEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT	551
Qy	491	TTKSAPTTKEPSPITTKPEPAPTTTKPEPAPTTTKKPEPAPTTTKPEPAPTTTKKP	550
Db	552	TTKSAPTTKEPSPITTKPEPAPTTTKKPEPAPTTTKKPEPAPTTTKPEPAPTTTKKP	611
Qy	551	APTAKPEPAPTTKEPAPTTKPEPAPTTKPEPAPTTTKEPAPTTTKEPAPTTTKEPAPT	610
Db	612	APTAKPEPAPTTKPEPAPTTKPEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT	671
Qy	611	PREPAPTTKAAAPNTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKGTAPTTLKEP	670
Db	672	PREPAPTTKAAAPNTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKGTAPTTLKEP	731
Qy	671	APTTPKAPKELAPTTTKEPSTTSKDPAPTTKGTAPTTKPEPAPTTKPEPAPTTKPG	730
Db	732	APTTPKAPKELAPTTTKEPSTTSKDPAPTTKGTAPTTKPEPAPTTKPEPAPTTKPG	791
Qy	731	TAPTILKEPAPTTKPKAPKELAPTTTKEPSTTSKDPAPTTKGTAPTTKPEPAPTTKPK	790
Db	792	TAPTILKEPAPTTKPKAPKELAPTTTKEPSTTSKDPAPTTKGTAPTTKPEPAPTTKPK	851
Qy	791	KPAPTTPETPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALENSPKPGVPT	850
Db	852	KPAPTTPETPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALENSPKPGVPT	911
Qy	851	TKTPTAKTPEMTTAKDKTERDLRTPETTTAAAPKMTKETAATTEKTESKITATTQV	910
Db	912	TKTPTAKTPEMTTAKDKTERDLRTPETTTAAAPKMTKETAATTEKTESKITATTQV	971
Qy	911	TSITTTQDTPPFKITTLTKTTLAPKVITTKITITTEIMNKPBEETAKPKDRATNSKATPK	970
Db	972	TSITTTQDTPPFKITTLTKTTLAPKVITTKITITTEIMNKPBEETAKPKDRATNSKATPK	1031
Qy	971	PQKPTAKPKPTSTKPKTMPRVKPKPTTTPRKMTSTMPELNPTSRIAEAMLQTTTRN	1030
Db	1032	PQKPTAKPKPTSTKPKTMPRVKPKPTTTPRKMTSTMPELNPTSRIAEAMLQTTTRN	1091
Qy	1031	QTPNSKLVEVNPXKSDAGAGETPHMLLRPHVFMPEVTPDMDYLPVNPQGIINPMLS	1090

Db	1092	QTPNSKLVEVNPXKSDAGAGETPHMLLRPHVFMPEVTPDMDYLPVNPQGIINPMLS	1151
RESULT 7			
ID	ADK65839	standard; protein; 1404 AA.	
XX	ADK65839;		
DT	06-MAY-2004	(first entry)	
XX	Angiogenesis-differentially expressed protein #53.		
DE	cytostatic; cardiant; vasotropic; antiarteriosclerotic;		
XX	angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;		
KW	gene expression; cancer; coronary artery disease; myocardial ischemia;		
KW	coronary arteriosclerosis; forensic medicine.		
OS	Homo sapiens.		
XX	WO2003066831-A2.		
PN	14-AUG-2003.		
PD	07-FEB-2003; 2003WO-US003848.		
XX	07-FEB-2002; 2002US-00067482.		
PR	10-JUN-2002; 2002US-00164595.		
PR	16-AUG-2002; 2002US-0403649P.		
PR	03-JAN-2003; 2003US-0437746P.		
XX	(ORIG-) ORIGENE TECHNOLOGIES INC.		
PA	Sun Z, Li X, Kovacs KP, Fan W, Jay G;		
PI	WPI; 2003-731502/69.		
XX	Determining the angiogenic index of a tissue or cell sample using		
XX	expression levels of differentially expressed genes, useful for		
PT	diagnosing or treating cancer, coronary artery disease, myocardial		
PT	ischemia and/or arteriosclerosis.		
XX	Disclosure; SEQ ID NO 78; 296pp; English.		
XX	The invention relates to a method of determining the angiogenic index of		
CC	a tissue or cell sample comprising assessing, in a sample, the expression		
CC	levels of one or more differentially-expressed gene from any of 34 DNA		
CC	sequences, given in the specification, where the levels are indicative of		
CC	the angiogenic index. The methods and compositions of the present		
CC	invention are useful for diagnosing, preventing and/or treating cancer,		
CC	coronary artery disease, myocardial ischemia or coronary		
CC	arteriosclerosis. They can also be used in research, drug discovery and		
CC	forensic medicine involving angiogenesis. This sequence corresponds to		
CC	one of the differentially expressed proteins of the invention.		
XX	Sequence 1404 AA;		
SQ	Query Match	99.3%; Score 5788; DB 7; Length 1404;	
	Best Local Similarity	95.4%; Pred. No. 3.1e-158;	
	Matches 1087; Conservative	0; Mismatches 3; Indels 50; Gaps 1;	
Qy	1	MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNOCHYMECCPDF	60
Db	1	MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNOCHYMECCPDF	60
Qy	61	KRVCTAELSCKGRCFESFERGECDCDAQCKKYDKCCPDYESFCA-----	105
Db	61	KRVCTAELSCKGRCFESFERGECDCDAQCKKYDKCCPDYESFCAEVNPTSPSSKKAP	120
Qy	106	-----EHSVSENQESSSSSSSSSSSTIW	130
Db	121	PPSGASQTIKSTTKRSPKPNKKTKVIESBEITEHVSSENQESSSSSSSSSTIR	180

QY 131 KIKSSKNSAANRELOKLVKDKKQNRKTKKPTPKPPVVDEAGSLDNGDFKVTTPDTS 190
 DB 181 KIKSSKNSAANRELOKLVKDKKQNRKTKKPTPKPPVVDEAGSLDNGDFKVTTPDTS 240
 QY 191 TQHNKYSTSPKLTAKPINRPSLSPNSDTSKETSITVKNKETTIVETKETTITNKQTSIDG 250
 DB 241 TQHNKYSTSPKLTAKPINRPSLSPNSDTSKETSITVKNKETTIVETKETTITNKQTSIDG 300
 QY 251 KEKTTSAKETQSIKTSKADLAPTQKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPAS 310
 DB 301 KEKTTSAKETQSIKTSKADLAPTQKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPAS 360
 QY 311 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 370
 DB 361 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 420
 QY 371 APTTTKSAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 430
 DB 421 APTTTKSAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 480
 QY 431 EPAPTAPKAPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 490
 DB 481 EPAPTAPKAPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 540
 QY 491 TTKSAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 550
 DB 541 TTKSAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 600
 QY 551 APTAPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 610
 DB 601 APTAPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 660
 QY 611 PEPAPTTPKAAPNTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 670
 DB 661 PEPAPTTPKAAPNTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 720
 QY 671 APTTPKBPAPKELAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 730
 DB 721 APTTPKBPAPKELAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 780
 QY 731 TAPTTLKBPAPTTTKBPAPKELAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 790
 DB 781 TAPTTLKBPAPTTTKBPAPKELAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 840
 QY 791 KPAPTTPTPTPTTSEVSTPTTKETPTTIHKSPDSTPELSAETPKALENSPKPGVPT 850
 DB 841 KPAPTTPTPTPTTSEVSTPTTKETPTTIHKSPDSTPELSAETPKALENSPKPGVPT 900
 QY 851 TKTPAATKPEMTTAKDKITRDLRTTPTTAAAPKMTKEATTTKETSKITATTQV 910
 DB 901 TKTPAATKPEMTTAKDKITRDLRTTPTTAAAPKMTKEATTTKETSKITATTQV 960
 QY 911 TSTTTQDPTTTPKLTITLTKVTTTKITITTEIMNKEETAKPKDRATNSKATTPK 970
 DB 961 TSTTTQDPTTTPKLTITLTKVTTTKITITTEIMNKEETAKPKDRATNSKATTPK 1020
 QY 971 FQKPTKAPKPTSTKPKTPMVRKPKTPTTPRKMSTMPNLNPTSRIAEAMLTQTTTPN 1030
 DB 1021 FQKPTKAPKPTSTKPKTPMVRKPKTPTTPRKMSTMPNLNPTSRIAEAMLTQTTTPN 1080
 QY 1031 QTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPMDVLPVPCQIILNPLS 1090
 DB 1081 QTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPMDVLPVPCQIILNPLS 1140

RESULT 8
 ADK67912
 ID ADK67912 standard; protein; 1311 AA.

XX ADK67912;

XX ADK67912;

DT 06-MAY-2004 (first entry)

DE Human extracellular messenger (EXMES) polypeptide.
 KW Human; extracellular messenger; EXMES; respiratory-gen.; antiallergic;
 KW antisthmatic; antiinflammatory; antidiabetic; neuroprotective;
 KW muscular-gen.; aniarthritic; osteopathic; hepatotropic; antipsoriatic;
 KW virucide; fungicide; antiparasitic; protozoacide; antihelminthic;
 KW cytostatic; gene therapy.
 XX Homo sapiens.
 OS
 XX
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..30
 FT /label= Signal_peptide
 FT /notes="Spans residues 1 to 18, 20, 21, 24, 29 or 30
 FT according to identification method"
 XX
 WO2004013292-A2.
 DD 12-FEB-2004.
 XX
 XX 30-JUL-2003; 2003WO-US024084.
 XX
 XX 02-AUG-2002; 2002US-0400810P.
 PR 19-SEP-2002; 2002US-0412197P.
 PR 04-OCT-2002; 2002US-0416004P.
 PR 08-NOV-2002; 2002US-0424862P.
 XX
 PA (INCY-) INCYTE CORP.
 XX
 XX Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP;
 PI Richardson TW, Emerling BM, Lindquist EA, Chawla NK, Ramkumar J;
 PI Lee SY;
 XX WPI; 2004-157116/15.
 DR N-PSDB; ADK67917.
 XX
 XX New extracellular messengers and nucleic acids, useful for diagnosing,
 PT treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes
 PT mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or
 PT autoimmune thyroiditis.
 XX
 PS Claim 60; SEQ ID NO 5; 165pp; English.
 XX
 CC The present sequence is that of novel human extracellular messenger
 CC (EXMES) Incyte ID NO: 7513018CD1 polypeptide. The protein shows homology
 CC to human megakaryocyte stimulating factor. The invention provides EXMES
 CC polynucleotides and polypeptides, as well as expression vectors, host
 CC cells, antibodies, agonists and antagonists, and methods for diagnosing,
 CC treating or preventing disorders associated with aberrant expression of
 CC EXMES, especially autoimmune and inflammatory disorders, cell
 CC proliferative disorders and endocrine disorders, e.g. adult respiratory
 CC distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's
 CC disease, diabetes mellitus, myasthenia gravis, osteoarthritis,
 CC osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal,
 CC parasitic, protozoal or helminthic infections, cancers, autoimmune
 CC thyroiditis, cretinism, Plummer's disease or thyroid carcinoma.
 CC Embodiments also provide methods for using the purified EXMES and/or
 CC their encoding polynucleotides for facilitating the drug discovery
 CC process, including determining of efficacy, dosage, toxicity and
 CC pharmacology, and for investigating the pathogenesis of diseases and
 CC medical conditions.
 XX
 SQ Sequence 1311 AA;

Query Match 96.3%; Score 5608.7; DB 8; Length 1311;
 Best Local Similarity 96.1%; Pred. No. 4, 1e-153;
 Matches 1047; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKLPYLLLLLLLVFIQQVSSQDLSSCAGRCGEGYSDATNCNDYHCYMECCPDF 60
 DB 1 MAWKLPYLLLLLLLVFIQQVSSQDLSSCAGRCGEGYSDATNCNDYHCYMECCPDF 60
 QY 61 KRVTAEILSCKGRCPESFERGECDCDAQCKYDKCCPDYEFCAAEHSVENQESSSS 120

Db 61 KRVTCTAELCKGRGCFESFERGECDDAQCCKYDKCCPDYBSFCAB----- 106
QY 121 SSSSSSTIWKIKSSKNKAANRELQCKLVKDNKNKRTKKKPTPKPPVVDVAGSLDNGD 180
Db 107 -----VKDNKNKRTKKKPTPKPPVVDVAGSLDNGD 137
QY 181 FKVTTPDSTTQHNKVSFSPKITTAKPINPRSLPNSDTSKETSITVKNKTTVETKETT 240
Db 138 FKVTTPDSTTQHNKVSFSPKITTAKPINPRSLPNSDTSKETSITVKNKTTVETKETT 197
QY 241 TTNKQSTDKGKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTKPEP 300
Db 198 TTNKQSTDKGKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTKPEP 257
QY 301 TTTTPKEPASTTKPEPTTIIKSAETTPKEPAPTTTKSAPITPKPEPAPTTKEPAPTPK 360
Db 258 TTTTPKEPASTTKPEPTTIIKSAETTPKEPAPTTTKSAPITPKPEPAPTTKEPAPTPK 317
QY 361 EPAPTTTKEPAPTTTKSAPITPKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPT 420
Db 318 EPAPTTTKEPAPTTTKSAPITPKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPT 377
QY 421 TKEPAPTTKEPAPTAAPKAPAPTTKEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPT 480
Db 378 TKEPAPTTKEPAPTAAPKAPAPTTKEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPT 437
QY 481 PTTTKEPAPTTTKSAPITPKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTPK 540
Db 438 PTTTKEPAPTTTKSAPITPKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTPK 497
QY 541 EPAPTTTKEPAPTAAPKAPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTPPEELAP 600
Db 498 EPAPTTTKEPAPTAAPKAPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTPPEELAP 557
QY 601 TPEEPTTTPKEPAPTTKAAAPNTPKPEPAPTTTKPEPAPTTTKPEPAPTTTKETAAPTPK 660
Db 558 TPEEPTTTPKEPAPTTKAAAPNTPKPEPAPTTTKPEPAPTTTKPEPAPTTTKETAAPTPK 617
QY 661 GTAPTTTKEPAPTTPKKAPKELAPTTTKEPASTTSDKPAPTTPKGTAPTTPKPEPAPTTP 720
Db 618 GTAPTTTKEPAPTTPKKAPKELAPTTTKEPASTTSDKPAPTTPKGTAPTTPKPEPAPTTP 677
QY 721 KEPAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPASTTSDKPAPTTPKGTAPTT 780
Db 678 KEPAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPASTTSDKPAPTTPKGTAPTT 737
QY 781 PKEPAPTTPKAPAPTTPEPPTTSEVSTPTTKEPPTTIHKSPDSTPELSAEPKALE 840
Db 738 PKEPAPTTPKAPAPTTPEPPTTSEVSTPTTKEPPTTIHKSPDSTPELSAEPKALE 797
QY 841 NSPKPFGVPTTKPAATKPEMTTAKDKTTERDLATTPETTTAAPKMTKETATTTKETE 900
Db 798 NSPKPFGVPTTKPAATKPEMTTAKDKTTERDLATTPETTTAAPKMTKETATTTKETE 857
QY 901 SKITATTTQVSTTQDTPPKITTLTKTTLAPKYTTTKTITTTIYNKKEETAKPKDR 960
Db 858 SKITATTTQVSTTQDTPPKITTLTKTTLAPKYTTTKTITTTIYNKKEETAKPKDR 917
QY 961 ATNSKATTPKOKPKAPKPKPTSTKKPKTMPVRKPKTTPTRKMTSTWPELNPTSRIAE 1020
Db 918 ATNSKATTPKOKPKAPKPKPTSTKKPKTMPVRKPKTTPTRKMTSTWPELNPTSRIAE 977
QY 1021 AMLQTTTRNQTPNSKLVENPKSEDAGGAEETHMILLRPHVFMPEVTPDMDYLPVFN 1080
Db 978 AMLQTTTRNQTPNSKLVENPKSEDAGGAEETHMILLRPHVFMPEVTPDMDYLPVFN 1037
QY 1081 QGIINPMLS 1090
Db 1038 QGIINPMLS 1047

RESULT 9

ADK65819
ID ADK65819 standard; protein; 1320 AA.
XX AC ADK65819;
XX DT 06-MAY-2004 (first entry)
XX DE Angiogenesis-differentially expressed protein ANH0316.
XX KW cytosolic; cardiant; vasotropic; antiarteriosclerotic;
XX KW angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
XX KW gene expression; cancer; coronary artery disease; myocardial ischemia;
XX KW coronary arteriosclerosis; forensic medicine.
OS Homo sapiens.
XX WO2003066831-A2.
XX PD 14-AUG-2003.
XX PF 07-FEB-2003; 2003WO-US003848.
XX PR 07-FEB-2002; 2002US-00067482.
XX PR 10-JUN-2002; 2002US-00164595.
XX PR 16-AUG-2002; 2002US-0403649P.
XX PR 03-JAN-2003; 2003US-0437746P.
XX (ORIG-) ORIGENE TECHNOLOGIES INC.
XX Sun Z, Li X, Kovacs KP, Fan W, Jay G;
XX WP; 2003-731502/69.
XX N-PSDB; ADK65818.
XX PT Determining the angiogenic index of a tissue or cell sample using
XX PT expression levels of differentially expressed genes useful for
XX PT diagnosing or treating cancer, coronary artery disease, myocardial
XX PT ischemia and/or arteriosclerosis.
XX PS Claim 23; SEQ ID NO 58; 296pp: English.
XX CC The invention relates to a method of determining the angiogenic index of
XX CC a tissue or cell sample comprising assessing, in a sample, the expression
XX CC levels of one or more differentially-expressed gene from any of 34 DNA
XX CC sequences, given in the specification, where the levels are indicative of
XX CC the angiogenic index. The methods and compositions of the present
XX CC invention are useful for diagnosing, preventing and/or treating cancer,
XX CC coronary artery disease, myocardial ischemia or coronary
XX CC arteriosclerosis. They can also be used in research, drug discovery and
XX CC forensic medicine involving angiogenesis. This sequence corresponds to
XX CC one of the differentially expressed proteins of the invention.

Sequence 1320 AA;

Query Match 92.2%; Score 5370.2; DB 7; Length 1320;
Best Local Similarity 92.3%; Pred. No. 3e-146;
Matches 1013; Conservative 8; Mismatches 28; Indels 48; Gaps 2;
QY 1 MAWKTPIYLLLSLVFVIQQVSSQDLSSCAGRGEGYSRDATCNDYCNQHYNECCPDF 60
Db 1 MAWKTPIYLLLSLVFVIQQVSSQ----- 25
QY 61 KRVTCTAELCKGRGCFESFERGECDDAQCCKYDKCCPDYBSFCABEHSVSENCSSSS 120
Db 26 -----ELSCRGRCFESFERGECDDAQCCKYDKCCPDYBSFCAEVHTSPSSSKAP 79
QY 121 SSSSSSTIWKIKSSKNKAANRELQCKLVKDNKNKRTKKKPTPKPPVVDVAG 173
Db 80 PPSGASQTIKSTTKRSPKPKPKKKTKKVIIESEITEVKNKNKRTKKKPTPKPPVVDVAG 139
QY 174 SGLDNGDFKVTTPDSTTQHNKVSFSPKITTAKPINPRSLPNSDTSKETSITVKNKETT 233
Db 140 SGLDNGDFKVTTPDSTTQHNKVSFSPKITTAKPINPRSLPNSDTSKETSITVKNKETT 199

QY 234 VETKETTTTNTKSTDDGKETTSAKETOSIEKTSKADLAPTSKVLAKETPKAETTTKGA 293
 Db 200 VETKETTTTNTKSTDDGKETTSAKETOSIEKTSKADLAPTSKVLAKETPKAETTTKGA 259
 QY 294 LTPKEPTTTTKEPASTTKEPTTTTTSAPTTPKEPAPTTTTSAPTTPKEPAPTTTKE 353
 Db 260 LTPKEPTTTTKEPASTTKEPTTTTTSAPTTPKEPAPTTTTSAPTTPKEPAPTTTKE 319
 QY 354 PAPTTPKEPAPTTTKEPAPTTTTSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 413
 Db 320 PAPTTPKEPAPTTTKEPAPTTTTSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 379
 QY 414 PKEPAPTTTKEPAPTTTKEPAPTTAPKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 473
 Db 380 PKEPAPTTTKEPAPTTTKEPAPTTAPKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 439
 QY 474 PTTTTSAPTTPKEPAPTTTTSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 533
 Db 440 PTTTTSAPTTPKEPAPTTTTSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 499
 QY 534 PAPTTPKEPAPTTTKEPAPTTAPKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 593
 Db 500 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 559
 QY 594 TPEELAPTTPEPTTTTPEEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 653
 Db 560 TPEELAPTTPEPTTTTPEEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 619
 QY 654 TAPTTPKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 713
 Db 620 TAPTTPKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 679
 QY 714 EPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 773
 Db 680 EPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 739
 QY 774 KETAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 833
 Db 740 KETAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 799
 QY 834 PTPKALENSPKPEGVPTTTPAATKPEMTTAKOKITRDLRTTPEITTAAPKMTKETAT 893
 Db 800 PTPKALENSPKPEGVPTTTPAATKPEMTTAKOKITRDLRTTPEITTAAPKMTKETAT 859
 QY 894 TTEKTESKITATTTQVSTTTQDTPPKITTLKTTTLAPKVTTKITTTTEIMNKPPEE 953
 Db 860 TTEKTESKITATTTQVSTTTQDTPPKITTLKTTTLAPKVTTKITTTTEIMNKPPEE 919
 QY 954 TAKPKDRATNSKATTPKQKPTKAPKPTSTKPKPTMVRKPKTTTPPRKWTSTMPDLN 1013
 Db 920 TAKPKDRATNSKATTPKQKPTKAPKPTSTKPKPTMVRKPKTTTPPRKWTSTMPDLN 979
 QY 1014 PFSRIAEAMLQTTTNPNTNSKLVNPKSEDAGAGETPHMLLRPHVMPVETDMD 1073
 Db 980 PFSRIAEAMLQTTTNPNTNSKLVNPKSEDAGAGETPHMLLRPHVMPVETDMD 1039
 QY 1074 YLPRVFNQGIINPMLS 1090
 Db 1040 YLPRVFNQGIINPMLS 1056

RESULT 10

ADK67911

ID ADK67911 standard; protein; 1270 AA.

XX AC ADK67911;

XX DT 06-MAY-2004 (first entry)

XX DE Human extracellular messenger (EXMES) polypeptide.

XX KW Human; extracellular messenger; EXMES; respiratory-gen.; antiallergic;

KW antiasthmatic; antiinflammatory; antiidiabetic; neuroprotective;
 KW muscular-gen.; antiarthritic; osteopathic; hepatotropic; antipsoriatic;
 KW virucide; fungicide; antiparasitic; protozoacide; antihelminthic;
 KW cytostatic; gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Peptide 1..29

XX PT /label= Signal peptide
 XX PT /notes= "Spans residues 1 to 18, 20, 21, 24, 27 or 29
 according to identification method"

XX WO2004013292-A2.

XX 12-FEB-2004.

XX 30-JUL-2003; 2003WO-US024084.

XX 02-AUG-2002; 2002US-0400810P.

XX 19-SEP-2002; 2002US-0412197P.

XX 04-OCT-2002; 2002US-0416004P.

XX 08-NOV-2002; 2002US-0424862P.

XX (INCY-) INCYTE CORP.

XX Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP;

XX Richardson TW, Emerling BM, Lindquist BA, Chawla NK, Ramkumar J;

XX Lee SY;

XX WPI; 2004-157116/15.

XX N-PSDB; ADK67916.

XX New extracellular messengers and nucleic acids, useful for diagnosing,
 PT treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes
 PT mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or
 PT autoimmune thyroiditis.

XX Claim 59; SEQ ID NO 4; 165pp: English.

XX The present sequence is that of novel human extracellular messenger
 CC (EXMES) incyte ID NO: 7513017CD1 polypeptide. The protein shows homology
 CC to human megakaryocyte stimulating factor. The invention provides EXMES
 CC polynucleotides and polypeptides, as well as expression vectors, host
 CC cells, antibodies, agonists and antagonists, and methods for diagnosing,
 CC treating or preventing disorders associated with aberrant expression of
 CC EXMES, especially autoimmune and inflammatory disorders, cell
 CC proliferative disorders and endocrine disorders, e.g. adult respiratory
 CC distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's
 CC disease, diabetes mellitus, myasthenia gravis, osteoarthritis,
 CC osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal,
 CC parasitic, protozoal or helminthic infections, cancers, autoimmune
 CC thyroiditis, cretinism, Plummer's disease or thyroid carcinoma.
 CC Embodiments also provide methods for using the purified EXMES and/or
 CC their encoding polynucleotides for facilitating the drug discovery
 CC process, including determining of efficacy, dosage, toxicity and
 CC pharmacology, and for investigating the pathogenesis of diseases and
 CC medical conditions.

XX Sequence 1270 AA;

Query Match 91.7%; Score 5344.6; DB 8; Length 1270;

Best Local Similarity 92.3%; Pred. No. 1.6e-145;

Matches 1006; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGCGEGSRDATCNCYNCHWECCEPDF 60

Db 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGCGEGSRDATCNCYNCHWECCEPDF 25

QY 61 KRVCTAELSCKRCFESFERGREGCDCAQCKKYDKCCPDYSEFCAEHSVENQSSSS 120

Db 26 -----ELSCKRCFESFERGREGCDCAQCKKYDKCCPDYSEFCAE----- 65

QY 121 SSSSSSIWKIKSSKNSAANRELOKLVKNDKKNRTKKKPTPKPPVVDVDSAGSLDNGD 180
Db 66 -----VKDNKKNRTKKKPTPKPPVVDVDSAGSLDNGD 96
QY 181 FKVTTPDSTTOHNVKVSPIKTTAKPINRPSIPNSDTSKETSIVNKKTETVETKETT 240
Db 97 FKVTTPDSTTOHNVKVSPIKTTAKPINRPSIPNSDTSKETSIVNKKTETVETKETT 156
QY 241 TTNKQTSIDGKKTISAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEP 300
Db 157 TTNKQTSIDGKKTISAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEP 216
QY 301 TPTTPEKPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTPK 360
Db 217 TPTTPEKPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTPK 276
QY 361 EPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 420
Db 277 EPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 336
QY 421 TKEPAPTTPKEPAPTAPKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSA 480
Db 337 TKEPAPTTPKEPAPTAPKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSA 396
QY 481 PTTTKEPAPTTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 540
Db 397 PTTTKEPAPTTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 456
QY 541 EPAPTTTKEPAPTAPKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPA 600
Db 457 EPAPTTTKEPAPTAPKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPA 516
QY 601 TPTTPEKPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTPK 660
Db 517 TPTTPEKPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTPK 576
QY 661 GTAPTTLKBPAPTTPKKPAKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTKEPA 720
Db 577 GTAPTTLKBPAPTTPKKPAKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTKEPA 636
QY 721 KBPAPTTPKGTAPTTLKBPAPTTPKKPAKELAPTTTKEPTSTTSKDPAPTTPKGT 780
Db 637 KBPAPTTPKGTAPTTLKBPAPTTPKKPAKELAPTTTKEPTSTTSKDPAPTTPKGT 696
QY 781 KBPAPTTPKGTAPTTLKBPAPTTPKKPAKELAPTTTKEPTSTTSKDPAPTTPKGT 840
Db 697 KBPAPTTPKGTAPTTLKBPAPTTPKKPAKELAPTTTKEPTSTTSKDPAPTTPKGT 756
QY 841 NSPKFPGVPTTKPAATKPEMTTAKDXTTERDLRTTPTTTAAAPKMTKETATTEKTE 900
Db 757 NSPKFPGVPTTKPAATKPEMTTAKDXTTERDLRTTPTTTAAAPKMTKETATTEKTE 816
QY 901 SKITATTQVTSSTTQDTPPKLITTLKWTLPAPKVTITTKKTTTTEIMNKPBEETAKPKDR 960
Db 817 SKITATTQVTSSTTQDTPPKLITTLKWTLPAPKVTITTKKTTTTEIMNKPBEETAKPKDR 876
QY 961 ATNSKATTPKPKAPKPKPTSKPKTMVPRVRKPKTTTPRKMSTSTMPNLNPTSRIAE 1020
Db 877 ATNSKATTPKPKAPKPKPTSKPKTMVPRVRKPKTTTPRKMSTSTMPNLNPTSRIAE 936
QY 1021 AMLOTTTRNQTPNSKLVNPKSDEAGAGETPHMLLRPHVFMPEVTPDMDYLPVRVN 1080
Db 937 AMLOTTTRNQTPNSKLVNPKSDEAGAGETPHMLLRPHVFMPEVTPDMDYLPVRVN 996
QY 1081 QGIIINPMLS 1090
Db 997 QGIIINPMLS 1006

RESULT 11
ABU53252
ID ABU53252 standard; protein; 546 AA.
XX

ABU53252;
14-APR-2003 (first entry)
Human testes-derived DKF2phtes3_4o19 homologue #1.
Human; gene therapy; vaccine; disease treatment; detection.
Homo sapiens.
WO200112659-A2.
22-FEB-2001.
18-AUG-2000; 2000WO-1B001496.
18-AUG-1999; 99US-0149499P.
28-SEP-1999; 99US-0156503P.
(GSHU-) GERMAN HUMAN GENOME PROJECT.
Wiemann S;
WPI; 2001-327840/34.
Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies.
Example III; Page 892; 1095pp; English.
This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence represents a homologue of a polypeptide described in the disclosure of the invention
Sequence 546 AA;
Query Match 50.1%; Score 2920; DB 4; Length 546;
Best Local Similarity 100.0%; Pred. No. 2.9e-76;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 438 KKPAPTTPKEPAPTTPKEPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPK 497
Db 1 KKPAPTTPKEPAPTTPKEPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPK 60
QY 498 TPKEPSPTTTKEPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPK 557
Db 61 TPKEPSPTTTKEPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPK 120
QY 558 PAPTTPKETATTPPKLITTLTPEKLAFTTPKPAFTTPEELAPTTPEPTTTEEPAPT 617
Db 121 PAPTTPKETATTPPKLITTLTPEKLAFTTPKPAFTTPEELAPTTPEPTTTEEPAPT 180
QY 618 TPKAAAPNTKEPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPK 677
Db 181 TPKAAAPNTKEPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPK 240
QY 678 PAPKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPKGPAPTTPKGPAPTTPKGT 737
Db 241 PAPKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPKGPAPTTPKGPAPTTPKGT 300
QY 738 EPAPTTPKKPAKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPKGTAPTTPKGT 797
Db 301 EPAPTTPKKPAKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPKGTAPTTPKGT 360
QY 798 ETPPTTSEVSTPTTKEPTTTHKSPDESTPELSAEPPTPKALENSPKPGVPTTKT 857

```

Db 361 ETPPTTSVSTPTTKETPTTIHKSPDESTPELSAEPPTKALENSPKBFGVETTKTPAAT 420
QY 858 KPEMTTAKDKTERDLRTTPTTAAAPKVTKETATTTTEKTTESKITATTTQVTSITTQD 917
Db 421 KPEMTTAKDKTERDLRTTPTTAAAPKVTKETATTTTEKTTESKITATTTQVTSITTQD 480
QY 918 TTPFKITTLTKTTLAPKVTTKKTIITTEIMNKPBETAKPKDRATNSKATTPKPKPTKA 977
Db 481 TTPFKITTLTKTTLAPKVTTKKTIITTEIMNKPBETAKPKDRATNSKATTPKPKPTKA 540
QY 978 PKKPTS 983
Db 541 PKKPTS 546

RESULT 12
ABU53253
ID ABU53253 standard; protein; 551 AA.
XX
AC ABU53253;
XX
DT 14-APR-2003 (first entry)
XX
DE Human testes-derived DKFzpthes3_4019 homologue #2.
XX
KW Human; gene therapy; vaccine; disease treatment; detection.
OS Homo sapiens.
XX
PN WO200112659-A2.
XX
PD 22-FEB-2001.
XX
PF 18-AUG-2000; 2000WO-IB001496.
XX
PR 18-AUG-1999; 99US-0149499P.
PR 28-SEP-1999; 99US-0156503P.
XX
PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX
PI Wiemann S;
XX
WPI; 2001-327840/34.
XX
Nucleic acids having the sequences of clones isolated from libraries of
different human tissues, useful in recombinant DNA methodologies.
XX
Example III; Page 892-893; 1095pp; English.
XX
This invention describes novel polynucleotides and polypeptides isolated
from human cDNA libraries which can be used for gene therapy or in
vaccines. The polynucleotides of the invention and antibodies encoded by
them may be used in the prevention, diagnosis and treatment of diseases
associated with inappropriate polypeptide expression. The products of the
invention may also be used to identify modulators of expression and
activity and to down regulate expression and activity. The antibodies of
the invention may also be used as diagnostic agents for detecting the
presence of polypeptides in samples. This sequence represents a homologue
of a polypeptide described in the disclosure of the invention
XX
SQ Sequence 551 AA;

Query Match 49.6%; Score 2888; DB 4; Length 551;
Best Local Similarity 100.0%; Pred. No. 2.4e-75;
Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 547 TKKPAPTAPKEPAPTTPKETAAPTTPKLTPTPEKLAAPTTPKEPAPTTPPELAAPTPEEP 606
Db 1 TKKPAPTAPKEPAPTTPKETAAPTTPKLTPTPEKLAAPTTPKEPAPTTPPELAAPTPEEP 60
QY 607 TPTTPEPAPTTPKAAAPNTKPEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKETAAPTTP 666
Db 61 TPTTPEPAPTTPKAAAPNTKPEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKETAAPTTP 120

```

```

QY 667 LKSPAPTTPKKPAPKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPKEPAPTTPKEPAPT 726
Db 121 LKSPAPTTPKKPAPKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPKEPAPTTPKEPAPT 180
QY 727 TPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPKEPAPT 786
Db 181 TPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPKEPAPT 240
QY 787 TTPKAPAPTTPETPTTSEVSTPTTTSKPTTIHKSPDESTPELSAEPPTKALENSPKEP 846
Db 241 TTPKAPAPTTPETPTTSEVSTPTTTSKPTTIHKSPDESTPELSAEPPTKALENSPKEP 300
QY 847 GVPTTKTPAATKPEMTTAKDKTERDLRTTPTTAAAPKVTKETATTTTEKTTESKITAT 906
Db 301 GVPTTKTPAATKPEMTTAKDKTERDLRTTPTTAAAPKVTKETATTTTEKTTESKITAT 360
QY 907 TQVTSITTQDTPFKITTLTKTTLAPKVTTKKTIITTEIMNKPBETAKPKDRATNSKA 966
Db 361 TQVTSITTQDTPFKITTLTKTTLAPKVTTKKTIITTEIMNKPBETAKPKDRATNSKA 420
QY 967 TTPKPKPTKAPKKPTSTKKPKTTPRVRKPKTTPRKMVTSTMPELNPTSRIAEAMLQTT 1026
Db 421 TTPKPKPTKAPKKPTSTKKPKTTPRVRKPKTTPRKMVTSTMPELNPTSRIAEAMLQTT 480
QY 1027 TRNQTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVPNOGIIN 1086
Db 481 TRNQTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVPNOGIIN 540
QY 1087 PMLS 1090
Db 541 PMLS 544

RESULT 13
AAB29778
ID AAB29778 standard; protein; 902 AA.
XX
AC AAB29778;
XX
DT 28-FEB-2001 (first entry)
XX
DE Human MSF-derived tribonectin.
XX
KW Human tribonectin; MSF; megakaryocyte stimulating factor;
alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
osteoarthritis; tribosupplementation; tissue adhesion inhibition;
friction coefficient reduction; gene therapy; antiarthritic; osteopathic.
XX
OS Homo sapiens.
XX
PN WO200064930-A2.
XX
PD 02-NOV-2000.
XX
PF 24-APR-2000; 2000WO-US010953.
XX
PR 23-APR-1999; 99US-00298970.
XX
PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
XX
PI Jay GD;
XX
WPI; 2001-024673/03.
XX
Novel tribonectin polypeptide useful as lubricant for treating
osteoarthritis, comprises O-linked lubricating moiety.
PT
PS Disclosure; Fig 1; 47pp; English.
XX
The invention relates to a human tribonectin which is a product of
alternative splicing of the human MSF (megakaryocyte stimulating factor)
gene. The tribonectin has at least one O-linked oligosaccharide
CC

```

CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats
 CC of a motif having at least 50% identity to the sequence KEPAPTT
 CC (AAB29774). The invention also relates to a nucleic acid encoding a human
 CC MSF-derived tribonectin; a biocompatible composition comprising a human
 CC tribonectin for inhibiting tissue adhesion formation; and a method of
 CC diagnosing osteoarthritis or a predisposition to osteoarthritis by
 CC measuring the amount of MSF or its fragment in a biological sample of a
 CC mammal, wherein an increased amount of MSF compared to a control
 CC indicates the presence of or predisposition to developing osteoarthritis.
 CC The tribonectin and DNA encoding it are useful in the treatment of
 CC osteoarthritis, where they may be used for lubricating mammalian joints,
 CC such as articulating joints of humans, dogs or horses. The tribonectin,
 CC when formulated as a membrane, foam, gel or fibre, is useful for
 CC inhibiting adhesion between two surfaces such as the injured tissues of a
 CC mammal, where the injury is caused by a surgical insertion or trauma, or
 CC an artificial device e.g., an orthopaedic implant. In particular, one of
 CC the surfaces is pericardial tissue. DNA encoding a tribonectin may be
 CC used in gene therapy. The present sequence represents a substantial
 CC portion of a human MSF-derived tribonectin
 CC
 XX Sequence 902 AA;
 SQ
 Query Match 49.4%; Score 2880.1; DB 4; Length 902;
 Best Local Similarity 71.4%; Pred. No. 8.1e-75;
 Matches 675; Conservative 27; Mismatches 74; Indels 169; Gaps 66;
 QY 150 VKDNKNRNTKKPTKPPVVDVDEAGSLDNGDFKVTTPDSTTQHNNKYSTSPKITTAKPIN 209
 Db 1 VKDNKNRNTKKPTKPPVVDVDEAGSLDNGDFKVTTPDSTTQHNNKYSTSPKITTAKPIN 60
 QY 210 PRPSLPNSDTSKETSLSLVNKEITVETKTTTNNKOTSDGKEKTTSAKETOSIEKTSNAK 269
 Db 61 PRPSLPNSDTSKETSLSLVNKEITVETKTTTNNKOTSDGKEKTTSAKETOSIEKTSNAK 120
 QY 270 DLAPTSKVLAKPTPKAETTTKGPALTTPKPEPTTPKPEASTTPKPEPTTTIKSAPTTPK 329
 Db 121 DLAPTSKVLAKPTPKAETTTKGPAL-----TTKPEASTTPKPEPTTTIKSAPTTPK 172
 QY 330 EPAPTTTKSNAPTTKEPAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPTT 389
 Db 173 EPAPTTTKSNAPTTKEPAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPTT 228
 QY 390 PKKPAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 449
 Db 229 -KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT 281
 QY 450 PPTPKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKSAPTTTKSAPTTTKPE 509
 Db 282 PTT-KEPAP-TTKPEAPTT-KEPAPTTKEPAP-TTKPEAPTTKEPAPTT-KEPAP-TTKE 335
 QY 510 PAPTTEKEPAPTTKPEAPTTKPEAPTTKPEAPTTTKKPAPTAPKPEAPTTTKETAPT 569
 Db 336 PAPTTEKEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT 388
 QY 570 TPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEPTTTTPEEPAPTTPKAAAPNTPKE 629
 Db 389 T-KEPAPTTKEP-APTKE-PAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KE 440
 QY 630 PAPTTEKEPAPTTKPEAPTTKETAAPTCKGATTLKPEAPTTKPEAPTT-KEPAPTT 687
 Db 441 PAPTTEKEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTTKEPAP-T 493
 QY 688 TKEPTSTSDKPAPTPKGTAPTTPKPEAPTTKPEAPTTTPKGTAPTTLKPEAPTTPKKP 747
 Db 494 TKEPAPTTKE-PAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEP 546
 QY 748 AP--KEPAPTTTKGPTSTSDKPAPTTPKETAPTTPKPEAPTTKPAPTTTPETPPPTS 805
 Db 547 APTTKEPAP-TTKPEAPTTKE-PAPTT-KEPAPTT-KEPAPTT-KEPAPTTKE-PAPTTK 600
 QY 806 EVSTPTTTKEPTTIHKSPPDSTSELSAEPTEKALENSKPEKPGVPTTKTAANKPEMTTFA 865
 Db 601 EPA--PTTKEPAPTTKEPAPTT-----KEPAP-----TTKEP-APTKEPAPTT----- 640

QY 866 KDKTTERDLRTTPETTTAAPKMTKETATTTEKTTESKITATTTOVTSTTODTTPFKITT 925
 Db 641 -----TKEPAPTT-----KEPAPTTKEPAPTTKEPAP----- 667
 QY 926 LKTTTLAPKVTTKKTTTTEIMNKPBEAKPKDRATNSKATTPKPKOKPKAPKPKPTSTK 985
 Db 668 -----TTKEPAPTTKEP-APT-TK 684
 QY 986 PKTMTPRVRKPKTTPTRKVTSTMPELNPTSRIAEAMLOTTTTRNQTNPNSKLVENVPKSE 1045
 Db 685 EP-----APTTPTRKVTSTMPELNPTSRIAEAML-TTTRNQTNPNSKLVENVPKSE 735
 QY 1046 DAGAEGETPHMLRPHVFPVETPDMYLPVENVQGIINPMLS 1090
 Db 736 DAGAEGETPHMLRPHVFPVETPDMYLPVENVQGIINPMLS 780
 RESULT 14
 ABUS3254
 ID ABUS3254 standard; protein; 513 AA.
 XX AC ABUS3254;
 XX DT 14-APR-2003 (first entry)
 XX DE Human testes-derived DRFZphes3_4019 homologue #3.
 XX KW Human; gene therapy; vaccine; disease treatment; detection.
 XX OS Homo sapiens.
 XX PN WO2000112659-A2.
 XX PD 22-FEB-2001.
 XX PF 18-AUG-2000; 2000WO-IB001496.
 XX PR 18-AUG-1999; 99US-0149499P.
 XX PR 28-SEP-1999; 99US-0156503P.
 XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
 XX PI Wiemann S;
 XX DR WPI; 2001-327840/34.
 XX PT Nucleic acids having the sequences of clones isolated from libraries of
 XX different human tissues, useful in recombinant DNA methodologies.
 XX PS Example III; Page 893; 1095pp; English.
 XX CC This invention describes novel polynucleotides and polypeptides isolated
 CC from human cDNA libraries which can be used for gene therapy or in
 CC vaccines. The polynucleotides of the invention and antibodies encoded by
 CC them may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate polypeptide expression. The products of the
 CC invention may also be used to identify modulators of expression and
 CC activity and to down regulate expression and activity. The antibodies of
 CC the invention may also be used as diagnostic agents for detecting the
 CC presence of polypeptides in samples. This sequence represents a homologue
 CC of a polypeptide described in the disclosure of the invention
 XX
 SQ Sequence 513 AA;
 Query Match 47.3%; Score 2757; DB 4; Length 513;
 Best Local Similarity 100.0%; Pred. No. 1.3e-71;
 Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 157 RTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDSTTQHNNKYSTSPKITTAKPINRPSLPP 216
 Db 1 RTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDSTTQHNNKYSTSPKITTAKPINRPSLPP 60

```

QY 217 NSDTSKETSILVNNKETTIVETKETTITNNKQTSIDGKEKTSKETSIAKTSIAKDLAPTSK 276
DB 61 NSDTSKETSILVNNKETTIVETKETTITNNKQTSIDGKEKTSKETSIAKTSIAKDLAPTSK 120
QY 277 VLAKEPTPKAETTTKGPALTTKPEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTPTT 336
DB 121 VLAKEPTPKAETTTKGPALTTKPEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTPTT 180
QY 337 KSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 396
DB 181 KSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 240
QY 397 TPKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP 456
DB 241 TPKEPAPTTPKEPTPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP 300
QY 457 APTTKPEPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 516
DB 301 APTTKPEPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 360
QY 517 EPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 576
DB 361 EPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 420
QY 577 TPPEKLAPTTPEKPAPTTPPEELAPTTPEEPTPTTPPEEPAPTTPKAAAPNTPEKPAPTTPK 636
DB 421 TPPEKLAPTTPEKPAPTTPPEELAPTTPEEPTPTTPPEEPAPTTPKAAAPNTPEKPAPTTPK 480
QY 637 EPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 669
DB 481 EPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 513

RESULT 15
AAR80041
ID AAR80041 standard; protein; 452 AA.
XX
AC AAR80041;
XX
DT 25-MAR-2003 (revised)
DT 10-APR-1996 (first entry)
XX
DE Human megakaryocytopoietin protein.
XX
KW Human; megakaryocytopoietin; wheat germ agglutinin; heparin;
KW megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;
KW multipotential stem cell.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 393..396
FT /note= "unspecified amino acids"
FT Misc-difference 444..446
FT /note= "unspecified amino acids"
XX
PN WC9523861-A1.
XX
PD 08-SEP-1995.
XX
PF 06-MAR-1995; 95WO-CN000015.
XX
PR 04-MAR-1994; 94CN-00112066.
XX
PA (SHAN-) SHANGHAI BEITE BIOTECHNOLOGY CO LTD.
XX
PI Gu X, Han Z, Shen Q;
XX
PI PI
XX
DR WPI; 1995-320576/41.
XX
DR N-PSDB; AAT04546.
XX
XX New haematopoietic cell growth factor - used for treating
XX thrombocytopenia and hematocytopenia.
PT

```

```

XX
PS
XX
CC This sequence represents the human megakaryocytopoietin (MPO) protein.
CC This sequence was purified using a carrier which can couple wheat germ
CC agglutinin and heparin to separate MPO. Fragments of this sequence (see
CC AAR80039 and AAR80040) were used to produce the amplification primers
CC shown in AAT04544 and AAT04545. The fragments amplified by these primers
CC can then be used as probes to screen human cDNA libraries for MPO cDNA.
CC The MPO cDNA can then be inserted into a plasmid which is used to
CC transform cells to produce MPO. The MPO sequence is capable of promoting
CC colony formation of megakaryocytes, enlarging the size of megakaryocytes
CC and stimulating the proliferation of multipotential stem cells. The
CC factor may be used for treating thrombocytopenia and hematocytopenia. The
CC purification method can be used to isolate MPO from human urine or serum
CC of patients with aplastic anaemia, and from animal blood or urine by
CC radiation exposing the animals to induce aplastic anaemia. (Updated on 25
CC -MAR-2003 to correct PA field.)
XX
SQ Sequence 452 AA;
Query Match 34.2%; Score 1994.2; DB 2; Length 452;
Best Local Similarity 83.3%; Pred. No. 9.7e-50;
Matches 378; Conservative 11; Mismatches 47; Indels 18; Gaps 2;
QY 1 MAWKTLPIVILLLLSVFVIQQVSSQDLSSCAGCGEGYSRDATCNCYNCOHYMECCPDF 60
DB 1 MAWKTLPIVILLLLSVFVIQQVSSQDLSSCAGCGEGYSRDATCNCYNCOHYMECCPDF 60
QY 61 KRVTAEELSCGRCFESFERGECDCDAQCKYDKCCPDYESFCAEHSVSENQSSSSS 120
DB 61 KRVTAEELSCGRCFESFERGECDCDAQCKYDKCCPDYESFCAEHSVSENQSSSSS 120
QY 121 SSSSSSTIWKIKSSKNSAANRELQKL-----KVQDNKKNRKTKKPTPKPPVVDAG 173
DB 121 PPSGASQTIKSTTKRSPKPNKKTKKVIESEITEVDKDNKKNRKTKKPTPKPPVVDAG 180
QY 174 SGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINRPSLPNSDTSKETSILVNNKETT 233
DB 181 SGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINRPSLPNSDTSKETSILVNNKETT 240
QY 234 VETKETTNNKQTSIDGKEKTSKETSIAKTSIAKDLAPTSKVLAKTPKAEITTKGPA 293
DB 241 VETKETTNNKQTSIDGKEKTSKETSIAKTSIAKDLAPTSKVLAKTPKAEITTKGPA 300
QY 294 LPTPKPEPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE 353
DB 301 LPTPKPEPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE 360
QY 354 PAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTPTT 413
DB 361 PAPTTPKEPAPTTPKEPAPTTPKSHPPPLPRSCXXXCTQP-----TPKEHPPL 409
QY 414 PKEPAPTTPKEPAPTTPKEPAPTAPKPAPTTPKE 447
DB 410 PRSLHPTTPKEPAPTTPKEPAPTAPKPAPLPLE 443

```

Search completed: October 13, 2004, 11:36:46
Job time : 93.8789 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 18.9474 Seconds
(without alignments)
3815.116 Million cell updates/sec

Title: SEQ1-B
Perfect score: 5826
Sequence: 1 MAWKTPYVLLLSVFIQ.....DMDYLPVNPQGIINPMLS 1090

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5826	100.0	1354	4	US-07-757-022B-48
2	5811	99.7	1140	4	US-07-757-022B-104
3	5811	99.7	1404	4	US-07-757-022B-2
4	5811	99.7	1404	4	US-07-757-022B-62
5	5811	99.7	1404	4	US-09-298-970A-1*
6	5788	99.3	1404	4	US-10-164-595-78
7	5643.3	96.9	1361	4	US-07-757-022B-40
8	5608.7	96.3	1311	4	US-07-757-022B-42
9	5571	95.6	1314	4	US-07-757-022B-50
10	5561.9	95.5	1049	4	US-07-757-022B-58
11	5561.9	95.5	1313	4	US-07-757-022B-142
12	5546.9	95.2	1363	4	US-07-757-022B-52
13	5486.7	93.8	1022	4	US-07-757-022B-84
14	5379.2	92.3	1320	4	US-07-757-022B-46
15	5379.2	92.3	1320	4	US-07-757-022B-60
16	5370.2	92.2	1320	4	US-10-164-595-58
17	5344.6	91.7	1038	4	US-07-757-022B-74
18	5344.6	91.7	1270	4	US-07-757-022B-44
19	5011	86.0	941	4	US-07-757-022B-14
20	1394.5	23.9	5179	4	US-09-538-092-1258
21	1091.9	18.7	8991	4	US-08-714-741-52
22	914	15.7	220	4	US-07-757-022B-96
23	904.3	15.5	207	4	US-07-757-022B-116
24	904.3	15.5	207	4	US-07-757-022B-136
25	869.7	14.9	157	4	US-07-757-022B-102
26	869.7	14.9	157	4	US-07-757-022B-114
27	852	14.6	209	4	US-07-757-022B-94

28	847	14.5	208	4	US-07-757-022B-132
29	838.2	14.4	463	4	US-07-757-022B-54
30	831.7	14.3	3256	4	US-09-919-172-98
31	831.7	14.3	3256	4	US-09-919-172-98
32	831.7	14.3	3256	4	US-09-919-039-21
33	826	14.2	204	4	US-07-757-022B-92
34	820.9	14.1	3118	3	US-09-579-181-1
35	813.6	14.0	2972	3	US-09-579-181-2
36	813.2	14.0	296	4	US-07-757-022B-70
37	801.6	13.8	231	4	US-07-757-022B-30
38	767	13.2	192	4	US-07-757-022B-90
39	764.4	13.1	1837	3	US-08-928-361B-5
40	764.4	13.1	1837	3	US-09-588-995A-5
41	761.8	13.1	2142	4	US-09-538-092-1142
42	748.7	12.9	132	4	US-07-757-022B-140
43	745.9	12.8	1721	3	US-08-700-551-5
44	745.9	12.8	1721	3	US-08-928-361B-6
45	745.9	12.8	1721	4	US-09-588-995A-6

ALIGNMENTS

RESULT 1
US-07-757-022B-48
; Sequence 48, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757.022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Coert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1354 amino acids
; TYPE: AMINO ACID

TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-48

Query Match 100.0%; Score 5826; DB 4; Length 1354;
Best Local Similarity 100.0%; Pred. No. 8.9e-172;
Matches 1090; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAWKTLPIVLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60
Db 1 MAWKTLPIVLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60

QY 61 KRVCTAELSCGRFCFSGRECDCAOCKYDKCCPDYEFCAEBHSVSNQSSSS 120
Db 61 KRVCTAELSCGRFCFSGRECDCAOCKYDKCCPDYEFCAEBHSVSNQSSSS 120

QY 121 SSSSSSTTWK1KSSXNSAANRELQKLVKONKORTKKKTPKPPVVDAGSLDNGD 180
Db 121 SSSSSSTTWK1KSSXNSAANRELQKLVKONKORTKKKTPKPPVVDAGSLDNGD 180

QY 181 FKVTTPDSTTQHNKYSTPKITAKPINRPSLPNSDTSKETSITVNETVETKTT 240
Db 181 FKVTTPDSTTQHNKYSTPKITAKPINRPSLPNSDTSKETSITVNETVETKTT 240

QY 241 TTNKQSTDGKKTTSKETSQTSIEKTSAXDLAPTSKVLAKTPKAEITTKGPALTTPKP 300
Db 241 TTNKQSTDGKKTTSKETSQTSIEKTSAXDLAPTSKVLAKTPKAEITTKGPALTTPKP 300

QY 301 TPTPKPEASTTPKEPTPTTIKSAPTTPKEPAPTTKSAPTTPKEPAPTTKPEAPTTPK 360
Db 301 TPTPKPEASTTPKEPTPTTIKSAPTTPKEPAPTTKSAPTTPKEPAPTTKPEAPTTPK 360

QY 361 EPAPTTTKPEAPTTKSAPTTPKEPAPTTKSAPTTPKEPAPTTKSAPTTPKEPAPTTK 420
Db 361 EPAPTTTKPEAPTTKSAPTTPKEPAPTTKSAPTTPKEPAPTTKSAPTTPKEPAPTTK 420

QY 421 TKEPAPTTKEPAPTPAKPAPKAPTTKEPAPTTKSAPTTPKEPAPTTKSAPTTPKEPAPTTK 480
Db 421 TKEPAPTTKEPAPTPAKPAPKAPTTKEPAPTTKSAPTTPKEPAPTTKSAPTTPKEPAPTTK 480

QY 481 PTTTKEPAPTTKSAPTTPKEPAPTTKSAPTTPKEPAPTTKSAPTTPKEPAPTTK 540
Db 481 PTTTKEPAPTTKSAPTTPKEPAPTTKSAPTTPKEPAPTTKSAPTTPKEPAPTTK 540

QY 541 EPAPTTTKPEAPTPAKPAPKAPTTKEPAPTTKSAPTTPKEPAPTTKSAPTTPKEPAPTTK 600
Db 541 EPAPTTTKPEAPTPAKPAPKAPTTKEPAPTTKSAPTTPKEPAPTTKSAPTTPKEPAPTTK 600

QY 601 TTPPEPTPTTPEPAPTTKAAAPNTPKAPAPTTKEPAPTTKSAPTTPKEPAPTTK 660
Db 601 TTPPEPTPTTPEPAPTTKAAAPNTPKAPAPTTKEPAPTTKSAPTTPKEPAPTTK 660

QY 661 GTAPTLKEPAPTTKAPKAPKAPLAPTTKEPAPTTKSAPTTPKEPAPTTKSAPTTPK 720
Db 661 GTAPTLKEPAPTTKAPKAPKAPLAPTTKEPAPTTKSAPTTPKEPAPTTKSAPTTPK 720

QY 721 KEAPAPTTKGTAPTLKEPAPTTKAPKAPKAPLAPTTKEPAPTTKSAPTTPKEPAPTTK 780
Db 721 KEAPAPTTKGTAPTLKEPAPTTKAPKAPKAPLAPTTKEPAPTTKSAPTTPKEPAPTTK 780

QY 781 PKEAPAPTTKAPAPTTPEPTPTTSEVSTPTTKEPAPTTKSAPTTPKEPAPTTKSAPTTPK 840
Db 781 PKEAPAPTTKAPAPTTPEPTPTTSEVSTPTTKEPAPTTKSAPTTPKEPAPTTKSAPTTPK 840

QY 841 NSPKPEPVPTTKTAPATKPEMTTAKDKTTERDLATTPETTTAAPKMTKETATTEKTE 900
Db 841 NSPKPEPVPTTKTAPATKPEMTTAKDKTTERDLATTPETTTAAPKMTKETATTEKTE 900

QY 901 SKITATTTQVSTTTQDTPPKITLKTTLAPKVTITTKKIIITTEIMNKPEETAKPKDR 960
Db 901 SKITATTTQVSTTTQDTPPKITLKTTLAPKVTITTKKIIITTEIMNKPEETAKPKDR 960

QY 961 ATNSKATTPKPKAPKPTSTKPKTMPVRKPTTPRKMSTMPBLNPTSIAE 1020
Db 961 ATNSKATTPKPKAPKPTSTKPKTMPVRKPTTPRKMSTMPBLNPTSIAE 1020

RESULT 2
US-07-757-022B-104
; Sequence 104, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-104

Query Match 99.7%; Score 5811; DB 4; Length 1140;
Best Local Similarity 95.6%; Pred. No. 2.1e-171;
Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIVLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60
Db 1 MAWKTLPIVLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60

we

QY 61 KRVCTAELSCGRCFESFERGECDCDAQCKKYDKCCPDYESFCA----- 105
Db 61 KRVCTAELSCGRCFESFERGECDCDAQCKKYDKCCPDYESFCAEVHNFTSPSSKKAP 120
QY 106 -----EHSVSENQESSSSSSSSSSSSSIW 130
Db 121 PPSGASQTIKSTTKRSPKPNKKTKVKVLESBEITEHSVSENQESSSSSSSSSSSIW 180
QY 131 KIKSSKNSAANRELQKLVKDKNKNRTKKKTPKPPVVVDEAGSLDNGDFKVTTPDTST 190
Db 181 KIKSSKNSAANRELQKLVKDKNKNRTKKKTPKPPVVVDEAGSLDNGDFKVTTPDTST 240
QY 191 TOHNKVSTSPKLTAKPINRSLPNSDTSKTSLSLVNKTETVTKETTTNKOTSDG 250
Db 241 TOHNKVSTSPKLTAKPINRSLPNSDTSKTSLSLVNKTETVTKETTTNKOTSDG 300
QY 251 KEKTTSAKETQSIKTSADKLAPTSKVLAKPTPKAETTTKGPAITTPKEPTTTPKEPAS 310
Db 301 KEKTTSAKETQSIKTSADKLAPTSKVLAKPTPKAETTTKGPAITTPKEPTTTPKEPAS 360
QY 311 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 370
Db 361 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 420
QY 371 APTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 430
Db 421 APTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 480
QY 431 EPAPTAPKKPAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPT 490
Db 481 EPAPTAPKKPAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPT 540
QY 491 TTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKKP 550
Db 541 TTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKKP 600
QY 551 APTAPKEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 610
Db 601 APTAPKEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 660
QY 611 PEAPTTTKPAAAPTTPKEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 670
Db 661 PEAPTTTKPAAAPTTPKEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 720
QY 671 APTTPKKAPKELAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTPKG 730
Db 721 APTTPKKAPKELAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTPKG 780
QY 731 TAPTTLKBPAPTTPKKAPKELAPTTTKGPTSTTSDKAPTTTKGPTSTTSDKAPTTTKPE 790
Db 781 TAPTTLKBPAPTTPKKAPKELAPTTTKGPTSTTSDKAPTTTKGPTSTTSDKAPTTTKPE 840
QY 791 KPAPTTTPEPTTSEVSTPTTKPEPTTIHKSPDSESTPELSAETPKALENSKPEGVPT 850
Db 841 KPAPTTTPEPTTSEVSTPTTKPEPTTIHKSPDSESTPELSAETPKALENSKPEGVPT 900
QY 851 TKTPAATKPEMTTAKDITERDLRTTPETTTAAKPMTKETATTTKTESKITATTQV 910
Db 901 TKTPAATKPEMTTAKDITERDLRTTPETTTAAKPMTKETATTTKTESKITATTQV 960
QY 911 TSTTTQDTPPKITTLKTTLAPKVTTKKITTEIMNKPEEETAKPKDRATNSKATTPK 970
Db 961 TSTTTQDTPPKITTLKTTLAPKVTTKKITTEIMNKPEEETAKPKDRATNSKATTPK 1020
QY 971 POKTPKAPKPTSTKPKPTMVRPKPTTPTPKMTSTMPNLNPTSRFAEAMLOTTTRPN 1030
Db 1021 POKTPKAPKPTSTKPKPTMVRPKPTTPTPKMTSTMPNLNPTSRFAEAMLOTTTRPN 1080
QY 1031 QTPNSKLVEVNPKSDAGAGETPHMLLRPHVFNPEVTPDMVLPVFNQGIILNPMLS 1090
Db 1081 QTPNSKLVEVNPKSDAGAGETPHMLLRPHVFNPEVTPDMVLPVFNQGIILNPMLS 1140

RESULT 3
US-07-757-022B-2
; Sequence 2, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-2
Query Match 99.7%; Score 5811; DB 4; Length 1404;
Best Local Similarity 95.6%; Pred. No. 2.7e-171;
Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1;
QY 1 MAWKTPYVILLLLSVFVIQQVSSQDLSSCAGRCGEGYGRDATCNCNDYNCQHYMECCPDF 60
Db 1 MAWKTPYVILLLLSVFVIQQVSSQDLSSCAGRCGEGYGRDATCNCNDYNCQHYMECCPDF 60
QY 61 KRVCTAELSCGRCFESFERGECDCDAQCKKYDKCCPDYESFCA----- 105
Db 61 KRVCTAELSCGRCFESFERGECDCDAQCKKYDKCCPDYESFCAEVHNFTSPSSKKAP 120
QY 106 -----EHSVSENQESSSSSSSSSSSSSIW 130
Db 121 PPSGASQTIKSTTKRSPKPNKKTKVKVLESBEITEHSVSENQESSSSSSSSSSSIW 180
QY 131 KIKSSKNSAANRELQKLVKDKNKNRTKKKTPKPPVVVDEAGSLDNGDFKVTTPDTST 190

QY 1031 QTPNSKLVNPKSDAGAGETPHMLLRPHVFMPEVTPDMVYLPRVNOGIIINPMLS 1090
Db 1081 QTPNSKLVNPKSDAGAGETPHMLLRPHVFMPEVTPDMVYLPRVNOGIIINPMLS 1140

RESULT 6
US-10-164-595-78
; Sequence 78, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-78

Query Match 99.3%; Score 5788; DB 4; Length 1404;
Best Local Similarity 95.4%; Pred. No. 1.4e-170;
Matches 1087; Conservative 0; Mismatches 3; Indels 50; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQVSSODLSSCAGGEGYSRDATCNCYNOHYMECCPDF 60
Db 1 MAWKTLPIYLLLSVFIQVSSODLSSCAGGEGYSRDATCNCYNOHYMECCPDF 60

QY 61 KRVCTAELSCRCRCPESPERGECDCDAOCKYDKCCPDYSEFCA----- 105
Db 61 KRVCTAELSCRCRCPESPERGECDCDAOCKYDKCCPDYSEFCAEVNHNPTSPSSKAP 120

QY 106 -----EEHSVSENOESSSSSSSSSSSSSTIW 130
Db 121 PPSGASQTIKSTTKRSPXPKKTKKVIIESEEITEHSVSENOESSSSSSSSSTIR 180

QY 131 KIKSSKNAANRELQKLVKNDKKNRTKPKTPPPVWDEAGSLGNDGPKVTPDTST 190
Db 181 KIKSSKNAANRELQKLVKNDKKNRTKPKTPPPVWDEAGSLGNDGPKVTPDTST 240

QY 191 TQHNKVSIPKITTAKPINRPSLPNSDTSKETS LVNKEITVETKETTINKQTSIDG 250
Db 241 TQHNKVSIPKITTAKPINRPSLPNSDTSKETS LVNKEITVETKETTINKQTSIDG 300

QY 251 KEKTSKAKETOSIEKTSADLAPTSKVLAKPTPKAETTKGPAULTTPKEPTTPKPEPAS 310
Db 301 KEKTSKAKETOSIEKTSADLAPTSKVLAKPTPKAETTKGPAULTTPKEPTTPKPEPAS 360

QY 311 TTPKEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 370
Db 361 TTPKEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 420

QY 371 APTTTKSAPTTKPEAPTTPKKPAPTTPKPEAPTTTKPEPTTPKPEAPTTKPEAPTTPK 430
Db 421 APTTTKSAPTTKPEAPTTPKKPAPTTPKPEAPTTTKPEPTTPKPEAPTTKPEAPTTPK 480

QY 431 EPAPTAPKPAPTTKPEAPTTKPEAPTTTKESPTTPKPEAPTTTKSAPTTTKPEAPTT 490
Db 481 EPAPTAPKPAPTTKPEAPTTKPEAPTTTKESPTTPKPEAPTTTKSAPTTTKPEAPTT 540

QY 491 TTKSAPTTKPEPSPTTKPEAPTTPKKPAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKKP 550
Db 541 TTKSAPTTKPEPSPTTKPEAPTTPKKPAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKKP 600

QY 551 APTAPKPEAPTTKPEATPTPKKLTPTTPEKLAETTPKPAPTTPBELAPTTPEEPTPTT 610
Db 601 APTTPKPEAPTTKPEATPTPKKLTPTTPEKLAETTPKPAPTTPBELAPTTPEEPTPTT 660

QY 611 PEEPAATPKAAAPNTPKEAPTTKPEAPTTTKPEAPTTTKETAPTTPKGTAPTTTLKEP 670

RESULT 7

US-07-757-022B-40
; Sequence 40, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseert, Luann
; REGISTRATION NUMBER: 31,822


```
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1361 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-40

Query Match          96.9%; Score 5643.3; DB 4; Length 1361;
Best Local Similarity 96.3%; Pred No. 3.8e-166;
Matches 1056; Conservative 8; Mismatches 26; Indels 7; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCDYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCDYNCQHYMECCPDF 60

QY 61 KXVCTAELSCRCRCFESFERGECDCDAQCKYDKCCPDYSEFCAEHSVSENESSSS 120
DB 61 KXVCTAELSCRCRCFESFERGECDCDAQCKYDKCCPDYSEFCAEHNPSTPPSSKKAP 120

QY 121 SSSSSSSITWIKTKSSKNGAANRELQKL-----KVDKNKNRTKKKPTPKPPVVDAG 173
DB 121 PPSGASQIKSTKESPPNNKKTKKVIESIEITEVDKNKNRTKKKPTPKPPVVDAG 180

QY 174 SGLDNGDKVTPDSTTQHNKVSTPKITAKPINRPSLPNPSDTSKETSITVKNKETT 233
DB 181 SGLDNGDKVTPDSTTQHNKVSTPKITAKPINRPSLPNPSDTSKETSITVKNKETT 240

QY 234 VETKETTNTNKTSTDGKEKTSKETSIAKTSKADLAPTSKVLAKPTPKAETTTKGA 293
DB 241 VETKETTNTNKTSTDGKEKTSKETSIAKTSKADLAPTSKVLAKPTPKAETTTKGA 300

QY 294 LPTPKPTTTPKBPASTPKPTTTPKSAPTTPKBPATTPKSAPTTPKBPATTPK 353
DB 301 LPTPKPTTTPKBPASTPKPTTTPKSAPTTPKBPATTPKSAPTTPKBPATTPK 360

QY 354 PAPTTPKBPATTPKBPATTPKSAPTTPKBPATTPKBPATTPKBPATTPKBPATTP 413
DB 361 PAPTTPKBPATTPKBPATTPKSAPTTPKBPATTPKBPATTPKBPATTPKBPATTP 420

QY 414 PKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBP 473
DB 421 PKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBP 480

QY 474 PPTTKSAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 533
DB 481 PPTTKSAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 540

QY 534 PAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 593
DB 541 PAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 600

QY 594 TPEELAPTPPEPTTTPPEAPPTPKAAAPNTPKEPATTPKBPATTPKBPATTPK 653
DB 601 TPEELAPTPPEPTTTPPEAPPTPKAAAPNTPKEPATTPKBPATTPKBPATTPK 660

QY 654 TAPTTPKGPATTLKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 713
DB 661 TAPTTPKGPATTLKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 720

QY 714 EPAPTTPKBPATTPKGPATTLKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 773
DB 721 EPAPTTPKBPATTPKGPATTLKBPATTPKBPATTPKBPATTPKBPATTPKBPAT 780

QY 774 KETAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBP 833
DB 781 KETAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBP 840

QY 834 PTPKALENSKPEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTPETTTAAPTKE 893
```

RESULT 8

```
US-07-757-022B-42
; Sequence 42, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1361 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-42
```

```
DB 841 PTPKALENSKPEPGVPTTKTPAATKPEMTTTAKDKTTERDLRTPETTTAAPTKE 900
QY 894 TTEKTTESKITATTTQVTSITTTQDTTPFKITTLKTTTLAPKVTITTKITITTEIMNKPEE 953
DB 901 TTEKTTESKITATTTQVTSITTTQDTTPFKITTLKTTTLAPKVTITTKITITTEIMNKPEE 960
QY 954 TAKPKORATNSKATTPKPKQPTKAPKKPTSTKKPKTMPRVKPKTTPTRKMTSTWPELN 1013
DB 961 TAKPKORATNSKATTPKPKQPTKAPKKPTSTKKPKTMPRVKPKTTPTRKMTSTWPELN 1020
QY 1014 PTSRIAEAMLQTTTRPNQTPNSKLVNPKSEDPAGAGETPHMLLRPHVFMPEVTPDMD 1073
DB 1021 PTSRIAEAMLQTTTRPNQTPNSKLVNPKSEDPAGAGETPHMLLRPHVFMPEVTPDMD 1080
QY 1074 YLPRVNPQGIINPMLS 1090
DB 1081 YLPRVNPQGIINPMLS 1097
```

```
Query Match          96.3%; Score 5608.7; DB 4; Length 1311;
Best Local Similarity 96.1%; Pred. No. 4.2e-165;
Matches 1047; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRGEGYSRDATCNCYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRGEGYSRDATCNCYNCQHYMECCPDF 60

QY 61 KEVCTAELSCGRCPESPERGECDCDAQCKKYDKCCPDYEFSCAEBHSVSENCSSSS 120
DB 61 KEVCTAELSCGRCPESPERGECDCDAQCKKYDKCCPDYEFSCAEBHSVSENCSSSS 106

QY 121 SSSSSSTIWKIKSSKNSAANRELQKKLVKNDKNRTKKKPTPKPVVDAGSGLNGD 180
DB 107 -----VKDNKNRTKKKPTPKPVVDAGSGLNGD 137

QY 181 FKVITPDSTTQHNKYSTPKITTAKPINPRPSLPNSDTSKETSLSLWNKETTIVTKETT 240
DB 138 FKVITPDSTTQHNKYSTSPKITTAKPINPRPSLPNSDTSKETSLSLWNKETTIVTKETT 197

QY 241 TTNKOTSTDGKEKITSAKETOSIEKTSADLAPTSSKVLAKPTPKAETTTKGPALTTTKEP 300
DB 198 TTNKOTSTDGKEKITSAKETOSIEKTSADLAPTSSKVLAKPTPKAETTTKGPALTTTKEP 257

QY 301 TPTPKPEASTTKEPTPTTTSKAPTTPKEPAPTTTTSAPTTPKEPAPTTTKEPAPTTPK 360
DB 258 TPTPKPEASTTKEPTPTTTSKAPTTPKEPAPTTTTSAPTTPKEPAPTTTKEPAPTTPK 317

QY 361 EPAPTTTKEPAPTTTTSKAPTTPKEPAPTTPKKAPTTTKEPAPTTTKEPAPTTTKEPAPT 420
DB 318 EPAPTTTKEPAPTTTTSKAPTTPKEPAPTTPKKAPTTTKEPAPTTTKEPAPTTTKEPAPT 377

QY 421 TKEPAPTTTKEPAPTTAPKKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 480
DB 378 TKEPAPTTTKEPAPTTAPKKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 437

QY 481 PTTTKEPAPTTTTSKAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 540
DB 438 PTTTKEPAPTTTTSKAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 497

QY 541 EPAPTTTKEPAPTTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 600
DB 498 EPAPTTTKEPAPTTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 557

QY 601 TTPPEPTTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 660
DB 558 TTPPEPTTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 617

QY 661 GTAPTTLKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 720
DB 618 GTAPTTLKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 677

QY 721 KEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 780
DB 678 KEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 737

QY 781 PKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 840
DB 738 PKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 797

QY 841 NSPKPEGVPTTKTAAKPEMTTAKDKTTERDLRTTPEPTTAAAPKMTKETATTTTEKITE 900
DB 798 NSPKPEGVPTTKTAAKPEMTTAKDKTTERDLRTTPEPTTAAAPKMTKETATTTTEKITE 857

QY 901 SKITATTTQVSTTQDTPPKITTLKTTLAPKVTITTKITTTTEIMNKPEETAKPKOR 960
DB 858 SKITATTTQVSTTQDTPPKITTLKTTLAPKVTITTKITTTTEIMNKPEETAKPKOR 917

QY 961 ATNSKATTPKPKTKAPKPTSTKPKTMRVRKPKTTTTPRKMSTMPELNPTSRIAE 1020
DB 918 ATNSKATTPKPKTKAPKPTSTKPKTMRVRKPKTTTTPRKMSTMPELNPTSRIAE 977

1021 AMLQTTTRNPQNSKLVNPNKSESDAGGAGETPHMLLRPHVFNPEVTPDMDYLPVFN 1080
978 AMLQTTTRNPQNSKLVNPNKSESDAGGAGETPHMLLRPHVFNPEVTPDMDYLPVFN 1037
1081 QGIINPMLS 1090
1038 QGIINPMLS 1047

RESULT 9
US-07-757-022B-50
; Sequence 50, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1314 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-50

Query Match          95.6%; Score 5571; DB 4; Length 1314;
Best Local Similarity 96.3%; Pred. No. 6e-164;
Matches 1050; Conservative 0; Mismatches 0; Indels 40; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRGEGYSRDATCNCYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRGEGYSRDATCNCYNCQHYMECCPDF 60

QY 61 KEVCTAELSCGRCPESPERGECDCDAQCKKYDKCCPDYEFSCAEBHSVSENCSSSS 120
DB 61 KEVCTAELSCGRCPESPERGECDCDAQCKKYDKCCPDYEFSCAEBHSVSENCSSSS 120
```

61	DB	KBVCTA-----EHSVSNQSSSSS	80
121	QY	SSSSSSSTTWIKSSKSNKAANRELQKLVKONKKNRTKKKPTPKPPVVDEAGSLDNGD	180
81	DB	SSSSSSSTTWIKSSKSNKAANRELQKLVKONKKNRTKKKPTPKPPVVDEAGSLDNGD	140
181	QY	FKVUTPDSTTGHNVKYSTSPKLTAKPINPRSLPNSDTSKETSIVNKEVTIKETT	240
141	DB	FKVUTPDSTTGHNVKYSTSPKLTAKPINPRSLPNSDTSKETSIVNKEVTIKETT	200
241	QY	TTNKQTSIDGKGKWTISAKETQSIKTSANDLAPTQKVLAKPTPKAETTTKGPALTTPKPEP	300
201	DB	TTNKQTSIDGKGKWTISAKETQSIKTSANDLAPTQKVLAKPTPKAETTTKGPALTTPKPEP	260
301	QY	TPPTPKPEPASTTPEBTPPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKPEPAPTTTKPEPAPTTK	360
261	DB	TPPTPKPEPASTTPEBTPPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKPEPAPTTTKPEPAPTTK	320
361	QY	EPAPTTTKPEPAPTTTKSAPTTPKEPAPTTPKPAPTTPEPAPTTPEPAPTTPEPAPTTPEPAPTTPEPAPTT	420
321	DB	EPAPTTTKPEPAPTTTKSAPTTPKEPAPTTPKPAPTTPEPAPTTPEPAPTTPEPAPTTPEPAPTTPEPAPTT	380
421	QY	TKPEAPTTTKPEPAPTAAPKAPAPTTPEPAPTTPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKSA	480
381	DB	TKPEAPTTTKPEPAPTAAPKAPAPTTPEPAPTTPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKSA	440
481	QY	PTTTTKPEAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTK	540
441	DB	PTTTTKPEAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTK	500
541	QY	EPAPTTTKKAPAPTAAPKAPAPTTPKETAPTTPKLPTTTPEKLAPTTTPEKLAPTTTPEKLAPTTTPEELAP	600
501	DB	EPAPTTTKKAPAPTAAPKAPAPTTPKETAPTTPKLPTTTPEKLAPTTTPEKLAPTTTPEKLAPTTTPEELAP	560
601	QY	TTTPEEBTPPTPEPAPTTTKAAAPNTPKPAPTTPEPAPTTPEPAPTTPEPAPTTPEPAPTTPEKETAATTPK	660
561	DB	TTTPEEBTPPTPEPAPTTTKAAAPNTPKPAPTTPEPAPTTPEPAPTTPEPAPTTPEPAPTTPEKETAATTPK	620
661	QY	GTAPTTTKPEPAPTTPKKAPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKPEPAPTTTP	720
621	DB	GTAPTTTKPEPAPTTPKKAPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTPKPEPAPTTTP	680
721	QY	KEPAPTTPKGTAPTTTKPEPAPTTPKKAPAPKELAPTTTKGTSTTSDKPAPTTPEKETAATTP	780
681	DB	KEPAPTTPKGTAPTTTKPEPAPTTPKKAPAPKELAPTTTKGTSTTSDKPAPTTPEKETAATTP	740
781	QY	PKPEPAPTTPKKAPAPTTPEPPTTSEVSTPTTKEPTTIHKSPDESTPELSAEPPTPKALE	840
741	DB	PKPEPAPTTPKKAPAPTTPEPPTTSEVSTPTTKEPTTIHKSPDESTPELSAEPPTPKALE	800
841	QY	NSPKPEGVPPTTKPAATKPMWTTAKDKTEROLRTTPEPPTTAAAPKMWKETATTTTEKTTE	900
801	DB	NSPKPEGVPPTTKPAATKPMWTTAKDKTEROLRTTPEPPTTAAAPKMWKETATTTTEKTTE	860
901	QY	SKITATTQVSTTTQDPTTFKLTTLKWTTLAPKVTTTTKTITTTTEIMNKPEETAKPKDR	960
861	DB	SKITATTQVSTTTQDPTTFKLTTLKWTTLAPKVTTTTKTITTTTEIMNKPEETAKPKDR	920
961	QY	ATNSKATTPPKQPKPTKAPKKPTSTTKPKTPRVKPKTTTPPKMTSTMPELNPTSGRIAE	1020
921	DB	ATNSKATTPPKQPKPTKAPKKPTSTTKPKTPRVKPKTTTPPKMTSTMPELNPTSGRIAE	980
1021	QY	AMLOTTTRPNQTPNSKLVFNPKSDEAGGAGETPHMLLEPHVFMPEVTPDMDYLPRVPN	1080
981	DB	AMLOTTTRPNQTPNSKLVFNPKSDEAGGAGETPHMLLEPHVFMPEVTPDMDYLPRVPN	1040
1081	QY	QGIHINPMLS 1090	
1041	DB	QGIHINPMLS 1050	
RESULT 10			
US-07-757-022B-58			

RESULT 10
US-07-757-

```

; Sequence 58, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1049 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-58

Query Match          95.5%; Score 5561.9; DB 4; Len
Best Local Similarity 96.2%; Pred. No. 8.6e-164;
Matches 1049; Conservative 0; Mismatches - 0; Indels - 0

QY      1 MAWKTPIVYLLLLSVFVYQVSSQDLSSCAGRCGGYSRDATCN
DB      1 MAWKTLPIYLLLLSVFVYQQVSSQ-----
QY      61 KRVTABLSCKGRCFESFGRGECDCDAQCCKYDKCCPDYVESFCA
DB      26 -----ELSCKGRCFESFGRGECDCDAQCCKYDKCCPDYVESFCA
QY      121 SSSSSSTIWKIKSSKNSAANELOKLVKDKNKQRTKKKPTPK
DB      80 SSSSSSTIWKIKSSKNSAANELOKLVKDKNKQRTKKKPTPK
QY      181 FKVITPDTSTQHNKVYSTSPKITTAKPINRPSPSPNSDTSKETS
DB      140 FKVITPDTSTQHNKVYSTSPKITTAKPINRPSPSPNSDTSKETS
QY      241 TTNKQSTGDKGKTTGAKETQSIKTSAXDLAPTQSKVLAKPTPK
DB      241 TTNKQSTGDKGKTTGAKETQSIKTSAXDLAPTQSKVLAKPTPK

```

Query Match 95.5%; Score 5561.9; DB 4; Length 1049;
Best Local Similarity 96.2%; Pred. No. 8.6e-164;

1	MAWKTLPIYLILLLSVFVIOQVSSQDILSSCAGRCGEGYSRDATNCNDYNCQHIMECCPDF	60
QY		
1	MAWKTLPIYLILLLSVFVIOQVSSQ-----	25
Db		
61	KRVCTABLSKCGRCFESPERGRECDCAQCKYDKCCPDYESFCAEHSYSVSENQESSSS	120
QY		
26	-----ELSKGRCFESPERGRECDCAQCKYDKCCPDYESFCAEHSYSVSENQESSSS	79
Db		
121	SSSSSSSTIWKISKSKSAANRELQKLVKDKNKQNTKKKPTPKPVVDEASGLDNGD	180
QY		
80	SSSSSSSTIWKISKSKSAANRELQKLVKDKNKQNTKKKPTPKPVVDEASGLDNGD	139
Db		
181	FKVITPPTDSTTQHNKVSTSPKITTAKPINRPSPPPNSDTSKETSLTVNKETTIVETKETT	240
QY		
140	FKVITPPTDSTTQHNKVSTSPKITTAKPINRPSPPPNSDTSKETSLTVNKETTIVETKETT	199
Db		
241	TTNKGTSTDGGKHTTSKAKETOSIKTSAXDLAPTSKVLAKPTPKAETTTKCPALITPKEP	300
QY		

```

Db 200 TTNKQSTGDKKTTSAKTSQIETSAKDLAPTSKVLAKPTPKAETTTKGPAITTPKEP 259
Qy 301 TPTTKEPASTTKEPTTTTISAKTTPKEPAPTTKSAPTTKEPAPTTTKEPAPTTK 360
Db 260 TPTTKEPASTTKEPTTTTISAKTTPKEPAPTTKSAPTTKEPAPTTTKEPAPTTK 319
Qy 361 EPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 420
Db 320 EPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 379
Qy 421 TKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 480
Db 380 TKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 439
Qy 481 PTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 540
Db 440 PTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 499
Qy 541 EPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 600
Db 500 EPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 559
Qy 601 TPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 660
Db 560 TPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 619
Qy 661 GTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 720
Db 620 GTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 679
Qy 721 KEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 780
Db 680 KEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 739
Qy 781 PKEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 840
Db 740 PKEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 799
Qy 841 NSPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 900
Db 800 NSPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 859
Qy 901 SKITATTTQVSTTTQDTTTPKTTTLLAPKVVTTTKEPTTTTKEPAPTTTKEPAPT 960
Db 860 SKITATTTQVSTTTQDTTTPKTTTLLAPKVVTTTKEPTTTTKEPAPTTTKEPAPT 919
Qy 961 ATNSKATTPKQKPTKAPKPTSTKPKTTPKVVVTTTKEPTTTTKEPAPTTTKEPAPT 1020
Db 920 ATNSKATTPKQKPTKAPKPTSTKPKTTPKVVVTTTKEPTTTTKEPAPTTTKEPAPT 979
Qy 1021 AMLQTTTTPNTPNSKLVNPKSDAGGAEGETPHMLLRPHVFMPEVTPDMOYLPRVFN 1080
Db 980 AMLQTTTTPNTPNSKLVNPKSDAGGAEGETPHMLLRPHVFMPEVTPDMOYLPRVFN 1039
Qy 1081 QGIIINPMLS 1090
Db 1040 QGIIINPMLS 1049

```

RESULT 11

```

US-07-757-022B-142
; Sequence 142, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.

```

```

; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-142

```

```

Query Match 95.5%; Score 5561.9; DB 4; Length 1313;
Best Local Similarity 96.2%; Pred. No. 1.2e-163;
Matches 1049; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

Qy 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25
Qy 61 KRVCATLSCGRCFESFERGECDCDAQCKKYDKCCPDYSCFAEHSVSENQSSSS 120
Db 26 -----ELSCGRCFESFERGECDCDAQCKKYDKCCPDYSCFAEHSVSENQSSSS 79
Qy 121 SSSSSSTIWKIKSKNSAANRELQKLKVXDNKKNRTKKKPTPKPPVVDAGSLDNGD 180
Db 80 SSSSSSTIWKIKSKNSAANRELQKLKVXDNKKNRTKKKPTPKPPVVDAGSLDNGD 139
Qy 181 FKVTTPTDSTTQHNKVTSPKITTAKPINRPSLPNPSDTSKETSIVNKETTVEKETT 240
Db 140 FKVTTPTDSTTQHNKVTSPKITTAKPINRPSLPNPSDTSKETSIVNKETTVEKETT 199
Qy 241 TTNKQSTGDKKTTSAKTSQIETSAKDLAPTSKVLAKPTPKAETTTKGPAITTPKEP 300
Db 200 TTNKQSTGDKKTTSAKTSQIETSAKDLAPTSKVLAKPTPKAETTTKGPAITTPKEP 259
Qy 301 TPTTKEPASTTKEPTTTTISAKTTPKEPAPTTKSAPTTKEPAPTTTKEPAPTTK 360
Db 260 TPTTKEPASTTKEPTTTTISAKTTPKEPAPTTKSAPTTKEPAPTTTKEPAPTTK 319
Qy 361 EPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 420
Db 320 EPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 379

```


Db 678 KPAPKELAPTTTKGPTSTSDKPAFTTTPKETAFTTTPKPAFTTTPKPAFTTTPETPTPTTS 737
QY 806 EVSTPTTKEPTTIHKSPDESTPELSAETTPKALENSPKPEPGVPTTKPAATKPEMTTITA 865
Db 738 EVSTPTTKEPTTIHKSPDESTPELSAETTPKALENSPKPEPGVPTTKPAATKPEMTTITA 797
QY 866 KOKTTERDLRTTPETTTAAPKMTKEATTTETKSTKITATTQVTSITTTQDTTPFKIIT 925
Db 798 KOKTTERDLRTTPETTTAAPKMTKEATTTETKSTKITATTQVTSITTTQDTTPFKIIT 857
QY 926 LKTTTLAPKVTITTKTITTEIMNKEETAKPKDRATNSKATTPKPKPTKAPKPTSTK 985
Db 858 LKTTTLAPKVTITTKTITTEIMNKEETAKPKDRATNSKATTPKPKPTKAPKPTSTK 917
QY 986 KPKTMPRVKPKPTTPPRKMTSTMPELNPTSIAEAMLOTTTRPQNTNSKLVEVNPKE 1045
Db 918 KPKTMPRVKPKPTTPPRKMTSTMPELNPTSIAEAMLOTTTRPQNTNSKLVEVNPKE 977
QY 1046 DAGGABGETPHMLLRHVFMPEVTPDMDYLPRVPMQIINPMLS 1090
Db 978 DAGGABGETPHMLLRHVFMPEVTPDMDYLPRVPMQIINPMLS 1022

RESULT 14
US-07-757-022B-46
; Sequence 46, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS: 143
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ceer, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: AMINO ACID

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-46
Query Match 92.3%; Score 5379.2; DB 4; Length 1320;
Best Local Similarity 92.5%; Pred. No. 4.9e-158;
Matches 1015; Conservative 8; Mismatches 26; Indels 48; Gaps 2;
QY 1 MAWTLPIYLLLSVFIQVSSQDLSSCAGRGEGYSRDATAICNDYNOHYMECCPDF 60
Db 1 MAWTLPIYLLLSVFIQVSSQ----- 25
QY 61 KRVCASLSCGRCFESFERGECDDAOCKKYDKCCPDYESFCABEHSYSENOESSSS 120
Db 26 -----ELSCGRCFESFERGECDDAOCKKYDKCCPDYESFCAEVHNTSPSSSKAP 79
QY 121 SSSSSSTIWKIKSSKNSAANRELQKL-----KVNKKQRTKKKPTPKPVDDEAG 173
Db 80 PPSGASQTIKSTTKRSPKPNKKTKKVI ESEITEVKDNKKQRTKKKPTPKPVDDEAG 139
QY 174 SGLDNGDFKVTTPDTSTTOHNVKYSTSKITATKPINRPSLPNSDTSKSTSLTVNKETT 233
Db 140 SGLDNGDFKVTTPDTSTTOHNVKYSTSKITATKPINRPSLPNSDTSKSTSLTVNKETT 199
QY 234 VETKETTITNKQISTDGKEXITTSKETSQSIKTSADLAPTSKVLAKPTKAEITTKGPA 293
Db 200 VETKETTITNKQISTDGKEXITTSKETSQSIKTSADLAPTSKVLAKPTKAEITTKGPA 259
QY 294 LTTPEKPTPTTPKBPASTTTPKEPTTTIKSAPTTKBPAPTTTTSASATTPKBPAPTTPKE 353
Db 260 LTTPEKPTPTTPKBPASTTTPKEPTTTIKSAPTTKBPAPTTTTSASATTPKBPAPTTPKE 319
QY 354 PAPTTPKEPAPTTTTPKBPATTTKSAPTTTPKEPAPTTTPKBPATTPKBPAPTTPKEPTPT 413
Db 320 PAPTTPKEPAPTTTTPKBPATTTKSAPTTTPKEPAPTTTPKBPATTPKBPAPTTPKEPTPT 379
QY 414 PKEPAPTTKBPAPTTTPKBPAPTA PKPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTPKE 473
Db 380 PKEPAPTTKBPAPTTTPKBPAPTA PKPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTPKE 439
QY 474 PTTTTSAPTTTTPKBPATTTKSAPTTTPKEPAPTTTPKBPAPTTTPKBPAPTTPKBPAPTTPKE 533
Db 440 PTTTTSAPTTTTPKBPATTTKSAPTTTPKEPAPTTTPKBPAPTTTPKBPAPTTPKBPAPTTPKE 499
QY 534 PAPTTPKEPAPTTTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAPTTPKBPAPTTPKE 593
Db 500 PAPTTPKEPAPTTTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPAPTTPKBPAPTTPKE 559
QY 594 TPBELAPTTPEEPTPTTPKEPAPTTTPKAAADNTPKBPAPTTTPKBPAPTTTPKBPAPTTPKE 653
Db 560 TPBELAPTTPEEPTPTTPKEPAPTTTPKAAADNTPKBPAPTTTPKBPAPTTTPKBPAPTTPKE 619
QY 654 TAPTTPKGTAPTTTPKBPATTPKBPAPKELAPTTTKEPTSTTSDKBPAPTTTPKGTAPTPK 713
Db 620 TAPTTPKGTAPTTTPKBPATTPKBPAPKELAPTTTKEPTSTTSDKBPAPTTTPKGTAPTPK 679
QY 714 EPAPTTPKEPAPTTTPKGTAPTTTPKBPAPTTTPKBPAPKELAPTTTKEPTSTTSDKBPAPTTP 773
Db 680 EPAPTTPKEPAPTTTPKGTAPTTTPKBPAPTTTPKBPAPKELAPTTTKEPTSTTSDKBPAPTTP 739
QY 774 KETAPTTTPKBPAPTTTPKBPAPTTTPETPTTSEVSTTTPKPTTTPKSTTSDKBPAPTTP 833
Db 740 KETAPTTTPKBPAPTTTPKBPAPTTTPETPTTSEVSTTTPKPTTTPKSTTSDKBPAPTTP 799
QY 834 PTPKALENSPKPEPGVPTTKTTPAATKPEMTTAKDKTERDLRTTPETTTAAPKMTKETAT 893
Db 800 PTPKALENSPKPEPGVPTTKTTPAATKPEMTTAKDKTERDLRTTPETTTAAPKMTKETAT 859
QY 894 TTEXTTESKITATTQVTSITTTQDTTPFKIITLTKTTLAPKVTITTKTITTEIMNKEE 953
Db 860 TTEXTTESKITATTQVTSITTTQDTTPFKIITLTKTTLAPKVTITTKTITTEIMNKEE 919
QY 954 TAKPKDRATNSKATTPKPKPTKAPKAPTSTKPKTMPRVKPKPTTPPKMTSTMPELN 1013

Db 920 TAKPKDRATNSKATTPKQKP:KAPKPTSTCKPKTMRVAKPKTTPPKWTSTMPDLN 979
QY 1014 PISRIAEAMLQTTTPNQTNSKLVENPKSDAGAGETPHMLLRPHVFMPEVTPDMD 1073
Db 980 PISRIAEAMLQTTTPNQTNSKLVENPKSDAGAGETPHMLLRPHVFMPEVTPDMD 1039
QY 1074 YLPRVFNQGIINPMLS 1090
Db 1040 YLPRVFNQGIINPMLS 1056

RESULT 15

US-07-757-022B-60
; Sequence 60, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseerr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-60

Query Match 92.3%; Score 5379.2; DB 4; Length 1320;
Best Local Similarity 92.5%; Pred. No. 4.9e-156;
Matches 1015; Conservative 8; Mismatches 26; Indels 48; Gaps 2;
QY 1 MAWKLTPIYLLLLLVFVIQVSSQDLSSCAGRCGEGYSDATNCNDYNCQHYMECCPDF 60
Db 1 MAWKLTPIYLLLLLVFVIQVSSQDLSSCAGRCGEGYSDATNCNDYNCQHYMECCPDF 25

QY 61 KRYCTAELSCKGRCPESFERGECDCDAOCKKYDKCCPDYESFCABEHSVSENQSSSS 120
Db 26 -----ELCKGRCPESFERGECDCDAOCKKYDKCCPDYESFCABEVNPTSPFSSKAP 79
QY 121 SSSSSSTTIWIKSSKNKSAANRELQKL-----XVDNKNKRTKKKTPKPPVVDKAG 173
Db 80 PPSGASQTIKSTTKRSPKPPNKKTKKVIESBEITEVDKNKKNRTKKKTPKPPVVDKAG 139
QY 174 SGLDNGDFKVTTPDSTTTOHNVSTSPKITTAKPINPRESLPPNSDTSKETSITVKNKETT 233
Db 140 SGLDNGDFKVTTPDSTTTOHNVSTSPKITTAKPINPRESLPPNSDTSKETSITVKNKETT 199
QY 234 VETKETTITNKQSTSDGKEKTTISAKETQSIETSAKDLPATSKVLAKPTPKAETTTKSPA 293
Db 200 VETKETTITNKQSTSDGKEKTTISAKETQSIETSAKDLPATSKVLAKPTPKAETTTKSPA 259
QY 294 LTTPKETPTTPKPEPASTTPKPTPTTIISAPTTKPEPAPTTTKSAPTTKPEPAPTTTKE 353
Db 260 LTTPKETPTTPKPEPASTTPKPTPTTIISAPTTKPEPAPTTTKSAPTTKPEPAPTTTKE 319
QY 354 PAPTTPKEPAPTTTKPEPAPTTTKSAPTTTPKPEPAPTTTKKPEPAPTTTPKPEPTPTT 413
Db 320 PAPTTPKEPAPTTTKPEPAPTTTKSAPTTTPKPEPAPTTTKKPEPAPTTTPKPEPTPTT 379
QY 414 PKPEPAPTTTPKPEPAPTTAPKAPADTTTPKPEPAPTTTPKPEPAPTTTPKPEPTTPKEPA 473
Db 380 PKPEPAPTTTPKPEPAPTTAPKAPADTTTPKPEPAPTTTPKPEPAPTTTPKPEPTTPKEPA 439
QY 474 PTTTKSAPTTTKPEPAPTTTKSAPTTTPKPEPAPTTTPKPEPAPTTTPKPEPAPTTTKE 533
Db 440 PTTTKSAPTTTKPEPAPTTTKSAPTTTPKPEPAPTTTPKPEPAPTTTPKPEPAPTTTKE 499
QY 534 PAPTTPKEPAPTTTKKAPAPAPKPEPAPTTKETAFTPKKLTPPTTPEKLAAPTTPKAPAPT 593
Db 500 PAPTTPKEPAPTTTKKAPAPAPKPEPAPTTKETAFTPKKLTPPTTPEKLAAPTTPKAPAPT 559
QY 594 TPPELAPTTPEPTPTTPPEPAPTTPKAAAPNTPKPEPAPTTTPKPEPAPTTTPKPEPAPTTKE 653
Db 560 TPPELAPTTPEPTPTTPPEPAPTTPKAAAPNTPKPEPAPTTTPKPEPAPTTTPKPEPAPTTTKE 619
QY 654 TAPTTPKGATPTTLKEPAPTTPKKAPAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPTTPK 713
Db 620 TAPTTPKGATPTTLKEPAPTTPKKAPAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPTTPK 679
QY 714 EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKAPAPKELAPTTTKEPTSTTSKAPAPTTP 773
Db 680 EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKAPAPKELAPTTTKEPTSTTSKAPAPTTP 739
QY 774 KETAPTTTPKEPAPTTTPKAPAPTTPETPPPTTSEVSTPTTKEPTTIHKSPDSESTPELSAE 833
Db 740 KETAPTTTPKEPAPTTTPKAPAPTTPETPPPTTSEVSTPTTKEPTTIHKSPDSESTPELSAE 799
QY 834 PTPKALENSPKPEGVPTTKTAAATKPEMTTAKDKTTERDLRTTTPETTTAAAPKMTKETAT 893
Db 800 PTPKALENSPKPEGVPTTKTAAATKPEMTTAKDKTTERDLRTTTPETTTAAAPKMTKETAT 859
QY 894 TTEKTTESKITATTTQVSTTTQDTPPKITTLTKTTLAPKVTITTKITTTTETIMNKPEE 953
Db 860 TTEKTTESKITATTTQVSTTTQDTPPKITTLTKTTLAPKVTITTKITTTTETIMNKPEE 919
QY 954 TAKPKDRATNSKATTPKQKPTKAPKPKPTSTKKKXTMPVRKPKTTPPKRMTSTNPELN 1013
Db 920 TAKPKDRATNSKATTPKQKPTKAPKPKPTSTKKKXTMPVRKPKTTPPKRMTSTNPELN 979
QY 1014 PISRIAEAMLQTTTPNQTNSKLVENPKSDAGAGETPHMLLRPHVFMPEVTPDMD 1073
Db 980 PISRIAEAMLQTTTPNQTNSKLVENPKSDAGAGETPHMLLRPHVFMPEVTPDMD 1039
QY 1074 YLPRVFNQGIINPMLS 1090
Db 1040 YLPRVFNQGIINPMLS 1056

Search completed: October 13, 2004, 11:58:50
Job time : 23.9474 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 110.895 Seconds
(without alignments)
3171.696 Million cell updates/sec

Title: SEQ1-B

Perfect score: 5826

Sequence: 1 MAWKTLPYLLLSVFVIQ.....DMDYLPVPGIINPMLS 1090

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1356558 seqs, 322682953 residues

Total number of hits satisfying chosen parameters: 1356558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

```
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5826	100.0	1354	13	US-10-124-557-48
2	5811	93.7	1140	13	US-10-124-557-104
3	5811	93.7	1404	9	US-09-802-207-30
4	5811	93.7	1404	11	US-09-897-188-1
5	5811	93.7	1404	13	US-10-124-557-2
6	5811	93.7	1404	13	US-10-124-557-62
7	5643.3	96.9	1361	13	US-10-124-557-40
8	5608.7	96.3	1311	13	US-10-124-557-42
9	5571	95.6	1314	13	US-10-124-557-50
10	5561.9	95.5	1049	13	US-10-124-557-58
11	5561.9	95.5	1313	13	US-10-124-557-142
12	5546.9	95.2	1363	13	US-10-124-557-52
13	5466.7	93.8	1022	13	US-10-124-557-84
14	5373.2	92.3	1320	13	US-10-124-557-46

```
15 5379.2 92.3 1320 13 US-10-124-557-60
16 5344.6 91.7 1038 13 US-10-124-557-74
17 5344.6 91.7 1270 13 US-10-124-557-44
18 5011 86.0 941 13 US-10-124-557-14
19 2446.8 42.0 732 9 US-09-802-207-27
20 1409.9 24.2 538 14 US-10-038-694-3
21 1394.5 23.9 5179 9 US-09-922-217-1068
22 1394.5 23.9 5179 9 US-09-833-263-1068
23 1394.5 23.9 5179 13 US-10-025-380-1068
24 1394.5 23.9 5179 16 US-10-734-564-111
25 1335.4 22.9 232 16 US-10-468-910-4
26 1010.7 17.3 188 14 US-10-038-694-2
27 960.8 16.5 1460 14 US-10-295-027-428
28 938.9 16.1 1325 9 US-09-864-761-35612
29 928.3 15.9 1367 9 US-09-801-368-108
30 914 15.7 220 13 US-10-124-557-96
31 904.3 15.5 207 13 US-10-124-557-116
32 904.3 15.5 207 13 US-10-124-557-136
33 901.5 15.5 3507 14 US-10-369-493-5784
34 869.7 14.9 157 13 US-10-124-557-102
35 869.7 14.9 157 13 US-10-124-557-114
36 862.5 14.8 6642 14 US-10-369-493-5013
37 859 14.7 2090 16 US-10-408-765A-2318
38 852 14.6 203 13 US-10-124-557-94
39 847 14.5 208 13 US-10-124-557-132
40 839.5 14.4 3256 16 US-10-408-765A-174
41 839.5 14.4 3256 16 US-10-701-490-9
42 838.2 14.4 463 13 US-10-124-557-54
43 836 14.3 19723 15 US-10-084-846A-5
44 832.7 14.3 697 15 US-10-425-114-41545
45 831.7 14.3 3256 9 US-09-919-172-98
```

ALIGNMENTS

RESULT 1

```
US-10-124-557-48
; Sequence 48, Application US/10124557
; Publication NO. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
;               Clark, Stephen C.
;               Jacobs, Kenneth C.
;               Hewick, Rodney M.
;               Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/10/124,557
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
```

```
Sequence 60, Appl
Sequence 74, Appl
Sequence 44, Appl
Sequence 14, Appl
Sequence 27, Appl
Sequence 3, Appl
Sequence 1068, Ap
Sequence 1068, Ap
Sequence 1068, Ap
Sequence 121, App
Sequence 4, Appl
Sequence 2, Appl
Sequence 428, App
Sequence 35612, A
Sequence 108, App
Sequence 96, Appl
Sequence 116, App
Sequence 136, App
Sequence 5784, Ap
Sequence 102, App
Sequence 114, App
Sequence 5013, Ap
Sequence 2318, Ap
Sequence 94, Appl
Sequence 132, App
Sequence 174, App
Sequence 9, Appl
Sequence 54, Appl
Sequence 5, Appl
Sequence 41545, A
Sequence 98, Appl
```



```
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104.

Query Match          99.7%; Score 5811; DB 13; Length 1140;
Best Local Similarity 95.6%; Pred. No. 5.1e-143;
Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60

QY 61 KEVCTAELSCKGRCFESFERGECDDCAQCKYDKCCPDYSEFCA----- 105
DB 61 KEVCTAELSCKGRCFESFERGECDDCAQCKYDKCCPDYSEFCA----- 105

QY 106 -----EHSVSENQESSSSSSSSSSSSSTI 130
DB 106 -----EHSVSENQESSSSSSSSSSSSSTI 130

QY 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEBITTEHSVSENQESSSSSSSSSTI 180
DB 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEBITTEHSVSENQESSSSSSSSSTI 180

QY 131 KIKSSKNSAANRELQKKLVKDNKNKRTKKKTPKPPVVDAGSLDNGDFKVTTPDST 190
DB 131 KIKSSKNSAANRELQKKLVKDNKNKRTKKKTPKPPVVDAGSLDNGDFKVTTPDST 190

QY 181 KIKSSKNSAANRELQKKLVKDNKNKRTKKKTPKPPVVDAGSLDNGDFKVTTPDST 240
DB 181 KIKSSKNSAANRELQKKLVKDNKNKRTKKKTPKPPVVDAGSLDNGDFKVTTPDST 240

QY 191 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTETTKTNNKQTSIDG 250
DB 191 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTETTKTNNKQTSIDG 250

QY 241 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTETTKTNNKQTSIDG 300
DB 241 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTETTKTNNKQTSIDG 300

QY 251 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 310
DB 251 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 310

QY 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 360
DB 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 360

QY 311 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 370
DB 311 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 370

QY 361 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 420
DB 361 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 420

QY 371 APTTTKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKGPALTTPK 430
DB 371 APTTTKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKGPALTTPK 430

QY 391 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 450
DB 391 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 450

QY 401 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 460
DB 401 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 460

QY 411 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 470
DB 411 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 470

QY 421 APTTTKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKGPALTTPK 480
DB 421 APTTTKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKGPALTTPK 480

QY 431 EPAPTAPKXPATTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKGPALTTPK 490
DB 431 EPAPTAPKXPATTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKGPALTTPK 490

QY 481 EPAPTAPKXPATTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKGPALTTPK 540
DB 481 EPAPTAPKXPATTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKGPALTTPK 540

QY 491 TTKSAPTTPKEPSPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTK 550
DB 491 TTKSAPTTPKEPSPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTK 550

QY 541 TTKSAPTTPKEPSPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTK 600
DB 541 TTKSAPTTPKEPSPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTK 600

QY 551 APTAPKEPATTPKETAATTPKLTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTK 610
DB 551 APTAPKEPATTPKETAATTPKLTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTK 610

QY 601 APTAPKEPATTPKETAATTPKLTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTK 660
DB 601 APTAPKEPATTPKETAATTPKLTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTK 660

QY 611 PPEPATTPKAAAPNTPKPATTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTK 670
DB 611 PPEPATTPKAAAPNTPKPATTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTK 670

QY 661 PPEPATTPKAAAPNTPKPATTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTK 720
DB 661 PPEPATTPKAAAPNTPKPATTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTK 720

QY 671 APTTPKAPKELAPTTTKPATTPKGPALTTPKEPATTTKGPALTTPKEPATTTKGPALTTPK 730
DB 671 APTTPKAPKELAPTTTKPATTPKGPALTTPKEPATTTKGPALTTPKEPATTTKGPALTTPK 730

QY 721 APTTPKAPKELAPTTTKPATTPKGPALTTPKEPATTTKGPALTTPKEPATTTKGPALTTPK 780
DB 721 APTTPKAPKELAPTTTKPATTPKGPALTTPKEPATTTKGPALTTPKEPATTTKGPALTTPK 780

QY 731 TAPTTLKBPATTPKAPKELAPTTTKPATTPKGPALTTPKEPATTTKGPALTTPKEPATTTK 790
DB 731 TAPTTLKBPATTPKAPKELAPTTTKPATTPKGPALTTPKEPATTTKGPALTTPKEPATTTK 790

QY 781 TAPTTLKBPATTPKAPKELAPTTTKPATTPKGPALTTPKEPATTTKGPALTTPKEPATTTK 840
DB 781 TAPTTLKBPATTPKAPKELAPTTTKPATTPKGPALTTPKEPATTTKGPALTTPKEPATTTK 840

QY 791 KPAPTTPPEPTPTTSEVSTPTTKPATTPKGPALTTPKEPATTTKGPALTTPKEPATTTK 850
DB 791 KPAPTTPPEPTPTTSEVSTPTTKPATTPKGPALTTPKEPATTTKGPALTTPKEPATTTK 850

QY 841 KPAPTTPPEPTPTTSEVSTPTTKPATTPKGPALTTPKEPATTTKGPALTTPKEPATTTK 900
DB 841 KPAPTTPPEPTPTTSEVSTPTTKPATTPKGPALTTPKEPATTTKGPALTTPKEPATTTK 900

QY 851 TKTPAATKEMTTAKOKTTERDLATTPETTTAAAPKMTKETATTTTEKTESKITATTQV 910
DB 851 TKTPAATKEMTTAKOKTTERDLATTPETTTAAAPKMTKETATTTTEKTESKITATTQV 910

QY 901 TKTPAATKEMTTAKOKTTERDLATTPETTTAAAPKMTKETATTTTEKTESKITATTQV 960
DB 901 TKTPAATKEMTTAKOKTTERDLATTPETTTAAAPKMTKETATTTTEKTESKITATTQV 960

QY 911 TSTTTQDTPPKITTLKTTTLAPKVTITTKKTTTTTEIMNKPETAKPKDRATNSKATTPK 970
DB 911 TSTTTQDTPPKITTLKTTTLAPKVTITTKKTTTTTEIMNKPETAKPKDRATNSKATTPK 970
```

```
DB 961 TSTTTQDTPPKITTLKTTTLAPKVTITTKKTTTTTEIMNKPETAKPKDRATNSKATTPK 1020
QY 971 PQKPTXAPKXPTSTKPKTMPRVKPKTTTPPKXMTSTMPKNTSTMPKNTSRIAEAMLQTTTRPN 1030
DB 1021 PQKPTXAPKXPTSTKPKTMPRVKPKTTTPPKXMTSTMPKNTSTMPKNTSRIAEAMLQTTTRPN 1080
QY 1031 QTSNSKLVEVNPXSSEDAGGAGETPHMLRPHVFMPEVTPDMVDYLPVFNQGIINPMLS 1090
DB 1081 QTSNSKLVEVNPXSSEDAGGAGETPHMLRPHVFMPEVTPDMVDYLPVFNQGIINPMLS 1140

RESULT 3
US-09-802-207-30
; Sequence 30, Application US/09802207
; Publication No. US2002008624A1
; GENERAL INFORMATION:
; APPLICANT: Warman, Matthew
; APPLICANT: Carpten, John
; APPLICANT: Trent, Jeffrey
; APPLICANT: Marcelino, Jose
; TITLE OF INVENTION: Novel Methods and Reagents for the Treatment of Osteoarthritis
; FILE REFERENCE: Case-06212
; CURRENT APPLICATION NUMBER: US/09/802,207
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 09/619,175
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/145,328
; PRIOR FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 30
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-802-207-30

Query Match          99.7%; Score 5811; DB 9; Length 1404;
Best Local Similarity 95.6%; Pred. No. 6.5e-143;
Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60

QY 61 KRVCATLSCKGRCFESFERGECDDCAQCKYDKCCPDYSEFCA----- 105
DB 61 KRVCATLSCKGRCFESFERGECDDCAQCKYDKCCPDYSEFCA----- 105

QY 106 -----BEHSVSENQESSSSSSSSSSSSSTI 130
DB 106 -----BEHSVSENQESSSSSSSSSSSSSTI 130

QY 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEBITTEHSVSENQESSSSSSSSSSSTI 180
DB 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEBITTEHSVSENQESSSSSSSSSSSTI 180

QY 131 KIKSSKNSAANRELQKKLVKDNKNKRTKKKTPKPPVVDAGSLDNGDFKVTTPDST 190
DB 131 KIKSSKNSAANRELQKKLVKDNKNKRTKKKTPKPPVVDAGSLDNGDFKVTTPDST 190

QY 181 KIKSSKNSAANRELQKKLVKDNKNKRTKKKTPKPPVVDAGSLDNGDFKVTTPDST 240
DB 181 KIKSSKNSAANRELQKKLVKDNKNKRTKKKTPKPPVVDAGSLDNGDFKVTTPDST 240

QY 191 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTETTKTNNKQTSIDG 250
DB 191 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTETTKTNNKQTSIDG 250

QY 241 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTETTKTNNKQTSIDG 300
DB 241 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTETTKTNNKQTSIDG 300

QY 251 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 310
DB 251 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 310

QY 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 360
DB 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAS 360

QY 311 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 370
DB 311 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 370

QY 361 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 420
DB 361 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 420

QY 371 APTTTKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKGPALTTPK 430
DB 371 APTTTKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKGPALTTPK 430

QY 391 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 450
DB 391 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 450

QY 401 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 460
DB 401 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 460

QY 411 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 470
DB 411 TTPKEPTPTTIKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKEP 470

QY 421 APTTTKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKGPALTTPK 480
DB 421 APTTTKSAPTTPKEPATTTKGPALTTPKEPATTTKGPALTTPKEPATTTKGPALTTPK 480
```


RESULT 5

US-10-124-557-2

; Sequence 2, Application US/10124557

; Publication No. US20020137894A1

GENERAL INFORMATION:

; APPLICANT: Turner, Katherine

; Clark, Stephen C.

; Jacobs, Kenneth

; Hewick, Rodney M.

; Gesner, Thomas G.

; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02140

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/124,557

; FILING DATE: 16-Apr-2002

; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/643,502

; FILING DATE: 18-JAN-1991

; APPLICATION NUMBER: US 07/546,114

; FILING DATE: 29-JUN-1990

; APPLICATION NUMBER: US 07/457,196

; FILING DATE: 29-DEC-1989

; APPLICATION NUMBER: US 07/390,901

; FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

; NAME: Cserr, Luann

; REGISTRATION NUMBER: 31,922

; REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617)876-1170

; TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

; LENGTH: 1404 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-10-124-557-2

Query Match 99.7%; Score 5811; DB 13; Length 1404;
 Best Local Similarity 95.6%; Pred. No. 6.5e-143;
 Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
 DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60

QY 61 KRVTAEALCKGRCFESFGRGECDDCAQCKYDKCCPDYESFCA----- 105
 DB 61 KRVTAEALCKGRCFESFGRGECDDCAQCKYDKCCPDYESFCAVHNFTSPPSKXAP 120

QY 106 -----EHSVSNQSSSSSSSSSSSSSSSIW 130
 DB 121 PPSGASQTIKSTTKRSPKPNKKTKVIESSEIIEHSVSNQSSSSSSSSSSSIW 180

QY 131 KIKSSKNSAANRELQKKLVKDNKNKRTKKKPTPKPPVVDEAGSLDNGDFKVTPTDST 190

RESULT 6

US-10-124-557-62

; Sequence 62, Application US/10124557

; Publication No. US20020137894A1

GENERAL INFORMATION:

; APPLICANT: Turner, Katherine

; Clark, Stephen C.

; Jacobs, Kenneth

; Hewick, Rodney M.

; Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 CambridgePark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557
 FILING DATE: 16-Apr-2002
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 62:

SEQUENCE CHARACTERISTICS:

LENGTH: 1404 amino acids
 TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 62:

US-10-124-557-62

Query Match 99.7%; Score 5811; DB 13; Length 1404;

Best Local Similarity 95.6%; Pred. No. 6.5e-143;

Matches 1090; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

1 MAWKTLPIYLLILLVSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60

1 MAWKTLPIYLLILLVSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60

61 KRVCCTAEISCKGRCPESFERGECDCDAQCKYDKCCPDYBSPCA----- 105

61 KRVCCTAEISCKGRCPESFERGECDCDAQCKYDKCCPDYBSCAEVHNFTSPSSKKAP 120

106 -----EEHVSSENQESSSSSSSSSSSTIW 130

121 PPSGASQIKTTKSPKPPNKKXKTVIESEEITEHVSSENQESSSSSSSSSSSTIW 180

131 KIKSKNSAANRELQKKLVKNDKNKNTKKTPTKPPVWDEAGSLDNGDFKVTPTST 190

181 KIKSKNSAANRELQKKLVKNDKNKNTKKTPTKPPVWDEAGSLDNGDFKVTPTST 240

191 TQHNKVSTSPKITTAKPINRPSLPNSDTSKESLTVNKETTVETKTTTNNQSTIDG 250

241 TQHNKVSTSPKITTAKPINRPSLPNSDTSKESLTVNKETTVETKTTTNNQSTIDG 300

251 KEKTSKETSISKTSKADLAPTSKVLAKPTKPAETTKGPAITTKPEPTTPKPEAS 310

301 KEKTSKETSISKTSKADLAPTSKVLAKPTKPAETTKGPAITTKPEPTTPKPEAS 360

311 TTPKEPTPTTIKSAPTTPKPAITTKSAPTTPKPAITTKSAPTTPKPAITTKKEPTTPK 370

361 TTPKEPTPTTIKSAPTTPKPAITTKSAPTTPKPAITTKSAPTTPKPAITTKKEPTTPK 420

371 APTTTKSAPTTPKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTKKEPTTPK 430

421 APTTTKSAPTTPKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTKKEPTTPK 480

431 EPAPTAPKPAITTPKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTKKEPTTPK 490

481 EPAPTAPKPAITTPKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTKKEPTTPK 540

491 TTKSAPTTPKPSPTTTKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTKK 550

541 TTKSAPTTPKPSPTTTKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTKK 600

551 APTAPKPAITTPKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTKPEEELAPTT 610

601 APTAPKPAITTPKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTKPEEELAPTT 660

611 PEEPAPTTPKAAAPNTKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTKK 670

661 PEEPAPTTPKAAAPNTKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTKK 720

671 APTTPKPAITTPKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTKK 730

721 APTTPKPAITTPKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTKK 780

731 TAPTTLKPAITTPKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTK 790

781 TAPTTLKPAITTPKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTK 840

791 KPAPTTPPETPTTSVSTPTTKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTK 850

841 KPAPTTPPETPTTSVSTPTTKPAITTKPAITTKPAITTKPAITTKPAITTKPAITTK 900

851 TKTPAATKPEMTTTAKDITKTERDLRTTPTTTTAAKWTKEATTTTETKTESKITATTQV 910

901 TKTPAATKPEMTTTAKDITKTERDLRTTPTTTTAAKWTKEATTTTETKTESKITATTQV 960

911 TSTTTQDPTTPPKITTLKTTLLAPKVTTKKITTITTEIMNKPEETAKPKDRATNSKATTPK 970

961 TSTTTQDPTTPPKITTLKTTLLAPKVTTKKITTITTEIMNKPEETAKPKDRATNSKATTPK 1020

971 POKPTKAPKPTSTKKEKTPRVRKPTTPRKTSTMPBLNPTSRABAMLOTTTPN 1030

1021 POKPTKAPKPTSTKKEKTPRVRKPTTPRKTSTMPBLNPTSRABAMLOTTTPN 1080

1031 QTPNSKLVNPKSEDAGAGETPHMLLRPHVFPVETPDMDYLPRVFNQGIINPMLS 1090

1081 QTPNSKLVNPKSEDAGAGETPHMLLRPHVFPVETPDMDYLPRVFNQGIINPMLS 1140

RESULT 7

US-10-124-557-40

; Sequence 40, Application US/10124557

; Publication No. US20020137894A1

; GENERAL INFORMATION:

; APPLICANT: Turner, Katherine

; Clark, Stephen C.

; Jacobs, Kenneth

; Hawick, Rodney M.

; Gesner, Thomas G.

; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESS: Genetics Institute, Inc.

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Juann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1361 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-124-557-40
Query Match 96.9%; Score 5643.3; DB 13; Length 1361;
Best Local Similarity 96.3%; Pred. No. 1.4e-138;
Matches 1056; Conservative 8; Mismatches 26; Indels 7; Gaps 1;
QY 1 MAWKLPYVILLVLLVVFVYVQVSSQDLSSCAGCGEGYSSDATCNCYNCOHYMECCPDF 60
DB 1 MAWKLPYVILLVLLVVFVYVQVSSQDLSSCAGCGEGYSSDATCNCYNCOHYMECCPDF 60
QY 61 KRVTAEKCKGRCFESFERGECDCDAQCKYDKCCPDYEFCAEHSVSSENQSSSS 120
DB 61 KRVTAEKCKGRCFESFERGECDCDAQCKYDKCCPDYEFCAEHSVSSENQSSSS 120
QY 121 SSSSSSTWIKSSKXNSAANRELQK-----KVONKQNRKTKKTPPPVVDK 173
DB 121 PPSGASQTIKSTTRGPKPPNKKTKVIESEBITEVKNKKNRTKKKTPPPVVDK 180
QY 174 SGLDNGDFKVTTPDTSTOHNVKYSTSPKITTAKPINRPSLPNSDTSKETSITVNNKETT 233
DB 181 SGLDNGDFKVTTPDTSTOHNVKYSTSPKITTAKPINRPSLPNSDTSKETSITVNNKETT 240
QY 234 VETKETTITNKQSTDGKETTSAKETQSIKTSADKLAPTSKVLAKTPKAEITTKGPA 293
DB 241 VETKETTITNKQSTDGKETTSAKETQSIKTSADKLAPTSKVLAKTPKAEITTKGPA 300
QY 294 LTPPKPPTTPPKPEASTTPKETPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKE 353
DB 301 LTPPKPPTTPPKPEASTTPKETPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKE 360
QY 354 PAPTTPKPEAPTTTKPEAPTTTKSAPTTPKPEAPTTTPKPEAPTTTPKPEPTPTT 413
DB 361 PAPTTPKPEAPTTTKPEAPTTTKSAPTTPKPEAPTTTPKPEAPTTTPKPEPTPTT 420
QY 414 PKPEAPTTKPEAPTTTPKPEAPTTAKKAPATTPKPEAPTTTPKPEAPTTTPKPEPTTKEPA 473
DB 421 PKPEAPTTKPEAPTTTPKPEAPTTAKKAPATTPKPEAPTTTPKPEAPTTTPKPEPTTKEPA 480
QY 474 PTTTKSAPTTKPEAPTTTKSAPTTPKPEPTTTTKPEAPTTTPKPEAPTTTPKPEPTTKE 533
DB 481 PTTTKSAPTTKPEAPTTTKSAPTTPKPEPTTTTKPEAPTTTPKPEAPTTTPKPEPTTKE 540
QY 534 PAPTTPKPEAPTTTKKPEAPTTAPKPEAPTTTPKETAPTTPKLTPPTPKLAPTTPEKPAPT 593

; FILING DATE: 29-DEC-1989
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cseerr, Luann
 ; REGISTRATION NUMBER: 31,822
 ; REFERENCE/DOCKET NUMBER: GI 5190
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)876-1170
 ; TELEFAX: (617)876-5851
 ; INFORMATION FOR SEQ ID NO: 42:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1311 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
 US-10-124-557-42

Query Match 96.3%; Score 5608.7; DB 13; Length 1311;
 Best Local Similarity 96.1%; Pred. No. 1.1e-137;
 Matches 1047; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY	1	MAWKTLPIYLLLSVFIQVSSODLSACGRCGEGYSRDATCNCYNCQHYMECCPDF	60
DB	1	MAWKTLPIYLLLSVFIQVSSODLSACGRCGEGYSRDATCNCYNCQHYMECCPDF	60
QY	61	KVCTAELSCGRCFESPERGECDCDAQCKYDKCCPDYSCAEHSVSENQESSSS	120
DB	61	KVCTAELSCGRCFESPERGECDCDAQCKYDKCCPDYSCAEHSVSENQESSSS	106
QY	121	SSSSSSSIWIKSSNSAANRELKLVKQDKNKQRTKKPTPKPPVVDAGSLONGD	180
DB	107	-----VKDNKNRTKKPTPKPPVVDAGSLONGD	137
QY	181	FKVTPDSTTQHNKVSPTKITTAKPINRPSLPNSDTSKETSITVNETTETK	240
DB	138	FKVTPDSTTQHNKVSPTKITTAKPINRPSLPNSDTSKETSITVNETTETK	197
QY	241	TNNKQTSDDGKEKITSKETOSIKTSKOLAPTSKVLAKPTPKAETTTKGPALTTPKEP	300
DB	198	TNNKQTSDDGKEKITSKETOSIKTSKOLAPTSKVLAKPTPKAETTTKGPALTTPKEP	257
QY	301	TPPTPKEPASTTPKEPTTTIKSAPTTKESAPTTTKSAPTTTKESAPTTTKESAPTTTK	360
DB	258	TPPTPKEPASTTPKEPTTTIKSAPTTKESAPTTTKSAPTTTKESAPTTTKESAPTTTK	317
QY	361	EPATTTKEPAPTTKSAPTTKESAPTTTKKAPTTTKESAPTTTKESAPTTTKESAPTTTK	420
DB	318	EPATTTKEPAPTTKSAPTTKESAPTTTKKAPTTTKESAPTTTKESAPTTTKESAPTTTK	377
QY	421	TKEPATTTKEPAPTAPEKAPTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTK	480
DB	378	TKEPATTTKEPAPTAPEKAPTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTK	437
QY	481	PTTTKEPAPTTKSAPTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTK	540
DB	438	PTTTKEPAPTTKSAPTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTK	497
QY	541	EPATTTTKKAPAPKAPKAPTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTK	600
DB	498	EPATTTTKKAPAPKAPKAPTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTK	557
QY	601	TTPEEPTTTPEEAPTTTKAANPNTPKEPAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTK	660
DB	558	TTPEEPTTTPEEAPTTTKAANPNTPKEPAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTK	617
QY	661	GTAPTTTKKAPAPKAPKAPTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTK	720
DB	618	GTAPTTTKKAPAPKAPKAPTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTK	677
QY	721	KEPAPTTPKGTAPTTKELAPTTKPKAPKAPKAPTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTK	780

DB	678	KEPAPTTPKGTAPTTKELAPTTKPKAPKAPKAPTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTKESAPTTTK	737
QY	781	PREPAPTTPKKAPPTTPEPTTSEVSTTTTKEPTTIHKSDESTPELSAETPKALE	840
DB	738	PREPAPTTPKKAPPTTPEPTTSEVSTTTTKEPTTIHKSDESTPELSAETPKALE	797
QY	841	NSPKPGVPTTKPAATKPEMTTAKDITERRDLRTTPEPTTAAPOKTKETATTTKTE	900
DB	798	NSPKPGVPTTKPAATKPEMTTAKDITERRDLRTTPEPTTAAPOKTKETATTTKTE	857
QY	901	SKITATTTQVSTTTTQDTPFKITLTKTTTLPKVTITTKTITTTIMNKPEETAKPKDR	960
DB	858	SKITATTTQVSTTTTQDTPFKITLTKTTTLPKVTITTKTITTTIMNKPEETAKPKDR	917
QY	961	ATNSKATTPKQPTKAPKPTSTKPKTMVPRVKPTTTPRKMSTWPELNPTSRIAE	1020
DB	918	ATNSKATTPKQPTKAPKPTSTKPKTMVPRVKPTTTPRKMSTWPELNPTSRIAE	977
QY	1021	AMLQTTTRPNQTPNSKLVNPKSEDAGGAGGETPHMLLRPHVEMPEVTPDMDYLPVRVN	1080
DB	978	AMLQTTTRPNQTPNSKLVNPKSEDAGGAGGETPHMLLRPHVEMPEVTPDMDYLPVRVN	1037
QY	1081	QGIINPMLS 1090	
DB	1038	QGIINPMLS 1047	

RESULT 9

US-10-124-557-50
 ; Sequence 50, Application US/10124557
 ; Publication No. US2002037894A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Turner, Katherine
 ; Jacobs, Kenneth
 ; Hewick, Rodney M.
 ; Gesner, Thomas G.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 Cambridgepark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION NUMBER: US/10/124,557
 ; FILING DATE: 16-Apr-2002
 ; CLASSIFICATION: <unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; APPLICATION NUMBER: US 07/546,114
 ; FILING DATE: 29-JUN-1990
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cseerr, Luann
 ; REGISTRATION NUMBER: 31,822
 ; REFERENCE/DOCKET NUMBER: GI 5190
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)876-1170
 ; TELEFAX: (617)876-5851
 ; INFORMATION FOR SEQ ID NO: 50:
 ; SEQUENCE CHARACTERISTICS:

```

;
; LENGTH: 1314 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-124-557-50

Query Match          95.6%; Score 5571; DB 13; Length 1314;
Best Local Similarity 96.3%; Pred. No. 1e-136;
Matches 1050; Conservative 0; Mismatches 0; Indels 40; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60
DB 1 MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60
QY 61 KXVCTAELSCGRCPESPERGECDCDAQCKYDKCCPDYESFCAEHSVSENOESSSS 120
DB 61 KXVCTA-----EHSVSENOESSSS 80
QY 121 SSSSSSTIWKTKSKNGAANRELQKLVKNDKNKRTKKKPTKPPVVDVAGSLDNGD 180
DB 81 SSSSSSTIWKTKSKNGAANRELQKLVKNDKNKRTKKKPTKPPVVDVAGSLDNGD 140
QY 181 FKVTTTPTDSTOHNVKSTPKITTAKPINRPSLPNSDTSKETSLSLVNKTIVETKTT 240
DB 141 FKVTTTPTDSTOHNVKSTPKITTAKPINRPSLPNSDTSKETSLSLVNKTIVETKTT 200
QY 241 ITNKQTSIDGKEKTSKQETQSIKTSKADLAPTSKVLAKPTPKAETTKGPAITTPKEP 300
DB 201 ITNKQTSIDGKEKTSKQETQSIKTSKADLAPTSKVLAKPTPKAETTKGPAITTPKEP 260
QY 301 TPTTKEPASTTPKEPTPTTIKSAPTTKEPAPTTTKSAPTTKEPAPTTTKEPAPTTPK 360
DB 261 TPTTKEPASTTPKEPTPTTIKSAPTTKEPAPTTTKSAPTTKEPAPTTTKEPAPTTPK 320
QY 361 EPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 420
DB 321 EPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 380
QY 421 TKEPAPTTTKEPAPTTAPKAPATTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 480
DB 381 TKEPAPTTTKEPAPTTAPKAPATTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 440
QY 481 PTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 540
DB 441 PTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 500
QY 541 EPAPTTTKEPAPTTAPKAPATTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 600
DB 501 EPAPTTTKEPAPTTAPKAPATTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 560
QY 601 TTPKEPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 660
DB 561 TTPKEPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 620
QY 661 GTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 720
DB 621 GTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 680
QY 721 KEAPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 780
DB 681 KEAPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 740
QY 781 PKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 840
DB 741 PKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 800
QY 841 NSPKFVGVTPTTAAKPEMTTAKDKTTERDLRTPTTPTTAAAPKMTKETATTEKTE 900
DB 801 NSPKFVGVTPTTAAKPEMTTAKDKTTERDLRTPTTPTTAAAPKMTKETATTEKTE 860
QY 901 SKITATTQVTSSTTQDTTTPFKITTLTKTTLAPKVTTKITTLTTEIMNKPEETAKPKDR 960
DB 901 SKITATTQVTSSTTQDTTTPFKITTLTKTTLAPKVTTKITTLTTEIMNKPEETAKPKDR 920
QY 961 ATNSKATTPKQKPTKAPKPKPTSTKKPKTMRVRKPKTTTTPRKMSTMTPELNPTSRIAE 1020
DB 921 ATNSKATTPKQKPTKAPKPKPTSTKKPKTMRVRKPKTTTTPRKMSTMTPELNPTSRIAE 980
QY 1021 AMLQTTTRNQTPNSKLVNPKSESDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPN 1080
DB 981 AMLQTTTRNQTPNSKLVNPKSESDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPN 1040
QY 1081 QGIIINPMLS 1090
DB 1041 QGIIINPMLS 1050

RESULT 10
US-10-124-557-58
; Sequence 58, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1049 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58

Query Match          95.5%; Score 5561.9; DB 13; Length 1049;
Best Local Similarity 96.2%; Pred. No. 1.3e-136;
Matches 1049; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60
```

Db 1 MAWKLPYVLLLSVFIQQVSSQ----- 25

QY 61 KRVCYTAELSCGRCFESFERGECDCDAQCKYDKCCPDYBSFCAEHSVSENQESSSS 120

Db 26 -----ELSCGRCFESFERGECDCDAQCKYDKCCPDYBSFCAEHSVSENQESSSS 79

QY 121 SSSSSSTIWKIKSSKNSAANRELQKLVKDNKNKRTKKKPTPKPPVVDAGSLDNGD 180

Db 80 SSSSSSTIWKIKSSKNSAANRELQKLVKDNKNKRTKKKPTPKPPVVDAGSLDNGD 139

QY 181 FKVTTPDSTTCHNKVSTSPKITTAKPINRPSLPNSDTSKETSLTVNKETTIVETKETT 240

Db 140 FKVTTPDSTTCHNKVSTSPKITTAKPINRPSLPNSDTSKETSLTVNKETTIVETKETT 199

QY 241 TTNKOTSDGKEKTTSAKETQSIKTSADLAPTSKVLAQTPKAEITTKGPAITTKPEP 300

Db 200 TTNKOTSDGKEKTTSAKETQSIKTSADLAPTSKVLAQTPKAEITTKGPAITTKPEP 259

QY 301 TPTTKPEASITTPKEPTTTIKSAPTTKPEAPITTKSAPTTKPEAPITTKPEAPITPK 360

Db 260 TPTTKPEASITTPKEPTTTIKSAPTTKPEAPITTKSAPTTKPEAPITTKPEAPITPK 319

QY 361 EPAPITTKPEAPITTKSAPTTKPEAPITTKPKAPITTKPEAPITTKPEAPITTKPEAPT 420

Db 320 EPAPITTKPEAPITTKSAPTTKPEAPITTKPKAPITTKPEAPITTKPEAPITTKPEAPT 379

QY 421 TKEAPITTKPEAPITTKPKAPITTKPEAPITTKPEAPITTKPEAPITTKPEAPITTKSA 480

Db 380 TKEAPITTKPEAPITTKPKAPITTKPEAPITTKPEAPITTKPEAPITTKPEAPITTKSA 439

QY 481 PTTTKPEAPITTKSAPTTKPEAPITTKPEAPITTKPKAPITTKPKAPITTKPKAPITTK 540

Db 440 PTTTKPEAPITTKSAPTTKPEAPITTKPEAPITTKPKAPITTKPKAPITTKPKAPITTK 499

QY 541 EPAPITTKPKAPITTKPEAPITTKPKAPITTKPKAPITTKPKAPITTKPKAPITTKPEELAP 600

Db 500 EPAPITTKPKAPITTKPEAPITTKPKAPITTKPKAPITTKPKAPITTKPKAPITTKPEELAP 559

QY 601 TTPPEPTTTTPEEAPITTKPKAAANTPKPEAPITTKPEAPITTKPEAPITTKPEAPITTK 660

Db 560 TTPPEPTTTTPEEAPITTKPKAAANTPKPEAPITTKPEAPITTKPEAPITTKPEAPITTK 619

QY 661 GTAPITTKPEAPITTKPKAPKELAPITTKPEPTTSDKAPITTKPKGAPITTKPEAPITTK 720

Db 620 GTAPITTKPEAPITTKPKAPKELAPITTKPEPTTSDKAPITTKPKGAPITTKPEAPITTK 679

QY 721 KEAPITTKGAPITTKPEAPITTKPKAPKELAPITTKGPTTSDKAPITTKPEAPITTKPEAPT 780

Db 680 KEAPITTKGAPITTKPEAPITTKPKAPKELAPITTKGPTTSDKAPITTKPEAPITTKPEAPT 739

QY 781 PREAPITTKPKAPITTKPEPTTSEVSTPTTKPEPTTIHKSPESTPELSAETPKALE 840

Db 740 PREAPITTKPKAPITTKPEPTTSEVSTPTTKPEPTTIHKSPESTPELSAETPKALE 799

QY 841 NSPKPEGVPITTKPAATKPEMTTAKOKTTRDRLRTPPTTAAAPKMTKETATTTETKTE 900

Db 800 NSPKPEGVPITTKPAATKPEMTTAKOKTTRDRLRTPPTTAAAPKMTKETATTTETKTE 859

QY 901 SKITATTTQVSTTTQDTPPKITTKTTLAPKVTITTKTITTEIMNKPEETAKPKDR 960

Db 860 SKITATTTQVSTTTQDTPPKITTKTTLAPKVTITTKTITTEIMNKPEETAKPKDR 919

QY 961 ATNKAITTKPKAPKKTSTTKPKTMPRVKPKTTPTRKMTSTMPELNPTSRIAE 1020

Db 920 ATNKAITTKPKAPKKTSTTKPKTMPRVKPKTTPTRKMTSTMPELNPTSRIAE 979

QY 1021 AMLQTTTTPNQTNSKLVNPKSDAGGAGETPHMLLRPHVPMPEVTPDMYLPRVNP 1080

Db 980 AMLQTTTTPNQTNSKLVNPKSDAGGAGETPHMLLRPHVPMPEVTPDMYLPRVNP 1039

QY 1081 QGIINPMLS 1090

Db 1040 QGIINPMLS 1049

RESULT 11

US-10-124-557-142

Sequence 142, Application US/10124557

Publication No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 876-1170

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 142:

SEQUENCE CHARACTERISTICS:

LENGTH: 1313 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 142:

US-10-124-557-142

Query Match 95.5%; Score 5561.9; DB 13; Length 1313;

Best Local Similarity 96.2%; Pred. No. 1.7e-136;

Matches 1049; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKLPYVLLLSVFIQQVSSQ----- 25

Db 1 MAWKLPYVLLLSVFIQQVSSQ----- 25

QY 61 KRVCYTAELSCGRCFESFERGECDCDAQCKYDKCCPDYBSFCAEHSVSENQESSSS 120

Db 26 -----ELSCGRCFESFERGECDCDAQCKYDKCCPDYBSFCAEHSVSENQESSSS 79

QY 121 SSSSSSTIWKIKSSKNSAANRELQKLVKDNKNKRTKKKPTPKPPVVDAGSLDNGD 180

Db 80 SSSSSSTIWKIKSSKNSAANRELQKLVKDNKNKRTKKKPTPKPPVVDAGSLDNGD 139

QY 181 FKVTTPDSTTCHNKVSTSPKITTAKPINRPSLPNSDTSKETSLTVNKETTIVETKETT 240

Db 140 FKVTTPDSTTCHNKVSTSPKITTAKPINRPSLSPNSDTSKETSLSLVNKETTIVTKETT 199
QY 241 TTNKQSTDSGKKTTSKETSQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEP 300
Db 200 TTNKQSTDSGKKTTSKETSQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEP 259
QY 301 TPTTPKEPASTTPKEPTPTTIKSAAPTTPKEPAPTTPKSAAPTTPKBPAPTTPKBPAPTTPK 360
Db 260 TPTTPKEPASTTPKEPTPTTIKSAAPTTPKEPAPTTPKSAAPTTPKBPAPTTPKBPAPTTPK 319
QY 361 EPAPTTTKEPAPTTPKSAAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 420
Db 320 EPAPTTTKEPAPTTPKSAAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 379
QY 421 TKEPAPTTPKEPAPTAPKAPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKSA 480
Db 380 TKEPAPTTPKEPAPTAPKAPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKSA 439
QY 481 PTTTKEPAPTTPKSAAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 540
Db 440 PTTTKEPAPTTPKSAAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 499
QY 541 EPAPTTTKEPAPTAPKAPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 600
Db 500 EPAPTTTKEPAPTAPKAPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 559
QY 601 TPEBPTTPPEPAPTTPKAAAPNTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 660
Db 560 TPEBPTTPPEPAPTTPKAAAPNTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 619
QY 661 GTAPTLTKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 720
Db 620 GTAPTLTKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 679
QY 721 KBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 780
Db 680 KBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 739
QY 781 KPEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 840
Db 740 KPEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 799
QY 841 NSPKBPAGVTTPKPAATKEMTTAKDKTTERDLRTTPTTTTAAKMTKETATTTKETE 900
Db 800 NSPKBPAGVTTPKPAATKEMTTAKDKTTERDLRTTPTTTTAAKMTKETATTTKETE 859
QY 901 SKITATTTQVSTTTQDTPPKITTLKTTTLAPKVTTKKTTTTEINMKPEETAKPKDR 960
Db 860 SKITATTTQVSTTTQDTPPKITTLKTTTLAPKVTTKKTTTTEINMKPEETAKPKDR 919
QY 961 ATNSKATTPKPKPTKAPKPTSTKPKTMPVRKPKTTPKMTSTMPBLNPTSRJAE 1020
Db 920 ATNSKATTPKPKPTKAPKPTSTKPKTMPVRKPKTTPKMTSTMPBLNPTSRJAE 979
QY 1021 AMLQTTTRENQTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVRVN 1080
Db 980 AMLQTTTRENQTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVRVN 1039
QY 1081 QGIINPMLS 1090
Db 1040 QGIINPMLS 1049

RESULT 12
US-10-124-557-52
; Sequence 52, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.

Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Csery, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 1363 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-124-557-52

Query Match 95.2%; Score 5546.9; DB 13; Length 1363;
Best Local Similarity 92.0%; Pred. No. 4.4e-136;
Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;
QY 1 MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSRDATAICNDYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLSLVFVIQVSSQ----- 25
QY 61 KRVCATLSCKGRCFESFERGREGDCDCAOCKKYDKCCPDYESFCA----- 105
Db 26 -----ELSCCKGRCFESFERGREGDCDCAOCKKYDKCCPDYESFCAEVHNFTSPSSKAP 79
QY 106 -----BEHSVSENOESSSSSSSSSSSSSSSTIW 130
Db 80 PPSGASQTIKSTTKESPKPNKKTKVIESEITEHSVSENOESSSSSSSSSSSSSSSTIW 139
QY 131 KIKSKNSAANRELQKKLVKDNKNKNTKKKTPKPPVDEAGSLDNGDFKVTTPDTST 190
Db 140 KIKSKNSAANRELQKKLVKDNKNKNTKKKTPKPPVDEAGSLDNGDFKVTTPDTST 199
QY 191 TQHNKVTSPKITTTAKPINRPSLPNSDTSKETSLSLVNKETTIVTKETTQSTIDG 250
Db 200 TQHNKVTSPKITTTAKPINRPSLPNSDTSKETSLSLVNKETTIVTKETTQSTIDG 259
QY 251 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPPKPEPAS 310
Db 260 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPPKPEPAS 319
QY 311 TTPKEPTTIKSAPTTPKBPAPTTPKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 370

DB 740 KETAPTPKEAPTPKPAETPTETPTPTTSEVSTPTTKEPTTIHKSPDESPELSAE 799
QY 834 PTPKALENSPKPEGVPTTKTAAATKPEMTTAKDKTTERDLRTTPEITTAAPKMTKETAT 893
DB 800 PTPKALENSPKPEGVPTTKTAAATKPEMTTAKDKTTERDLRTTPEITTAAPKMTKETAT 859
QY 894 TTEKTTESKITATTTQVSTTTQDTPPKIITTLKTIILAPKVTTKKTIITTEIMNKPEE 953
DB 860 TTEKTTESKITATTTQVSTTTQDTPPKIITTLKTIILAPKVTTKKTIITTEIMNKPEE 919
QY 954 TAKPKDRATNSKATTPKPKPTKAPKPKTSKKPKTWPVRVKPTTTPPKMTSTMPELN 1013
DB 920 TAKPKDRATNSKATTPKPKPTKAPKPKTSKKPKTWPVRVKPTTTPPKMTSTMPELN 979
QY 1014 PPSRIAEAMLQTTTPRNPOTPNKSLVEVNPKESEDAGAEGETPHMLLRPHVFMPEVTPDMD 1073
DB 980 PPSRIAEAMLQTTTPRNPOTPNKSLVEVNPKESEDAGAEGETPHMLLRPHVFMPEVTPDMD 1039
QY 1074 YLPRVNOGIIINPMLS 1090
DB 1040 YLPRVNOGIIINPMLS 1056

RESULT 15

US-10-124-557-60

Sequence 60, Application US/10124557

Publication No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

Geisner, Thomas G.

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-124-557-60
Query Match 92.3%; Score 5379.2; DB 13; Length 1320;
Best Local Similarity 92.5%; Pred. No. 9.5e-132;
Matches 1015; Conservative 8; Mismatches 26; Indels 48; Gaps 2;
QY 1 MANKTLPIYLLALLLVFVIQVSSQDLSSCAGRCGEGYSRQATCNDYNCQHYMECCPDF 60
DB 1 MANKTLPIYLLALLLVFVIQVSSQDLSSCAGRCGEGYSRQATCNDYNCQHYMECCPDF 25
QY 61 KRVCTAELCKGRCEPESFERGREGCDCAQCKYDKCCPDYSCFAEHSVSNSQSSSS 120
DB 26 -----ELSKGRCEPESFERGREGCDCAQCKYDKCCPDYSCFAEHSVSNSQSSSS 79
QY 121 SSSSSSTTWKTSKNSAANRELOKLL-----KVDNKNRTKKKPTKPPVVDGAG 173
DB 80 PPSGASQTIKSTTKSPKPFNKKTCKVISEBEITEVDNKNRTKKKPTKPPVVDGAG 139
QY 174 SGLDNGDFKVTTPDSTTOHNVSTSPKITTAKPINPRPSLPNSDTSKETSITVNEKIT 233
DB 140 SGLDNGDFKVTTPDSTTOHNVSTSPKITTAKPINPRPSLPNSDTSKETSITVNEKIT 199
QY 234 VETKETTNTNKTSTDGKEKTSKETSITVNEKIT-----KVDNKNRTKKKPTKPPVVDGAG 293
DB 200 VETKETTNTNKTSTDGKEKTSKETSITVNEKIT-----KVDNKNRTKKKPTKPPVVDGAG 259
QY 294 LITPKETPTTPKPEASTTPKPTPTTIKSAFTTPKPEASTTPKPEASTTPKPEASTTPK 353
DB 260 LITPKETPTTPKPEASTTPKPTPTTIKSAFTTPKPEASTTPKPEASTTPKPEASTTPK 319
QY 354 PAPTTPKPEASTTPKPEASTTPKPTPTTIKSAFTTPKPEASTTPKPEASTTPKPEASTTPK 413
DB 320 PAPTTPKPEASTTPKPEASTTPKPTPTTIKSAFTTPKPEASTTPKPEASTTPKPEASTTPK 379
QY 414 PKEPAPTTPKPEASTTPKPEASTTPKPTPTTIKSAFTTPKPEASTTPKPEASTTPKPEASTTPK 473
DB 380 PKEPAPTTPKPEASTTPKPEASTTPKPTPTTIKSAFTTPKPEASTTPKPEASTTPKPEASTTPK 439
QY 474 PTTTKSAPTTPKPEASTTPKPEASTTPKPTPTTIKSAFTTPKPEASTTPKPEASTTPKPEASTTPK 533
DB 440 PTTTKSAPTTPKPEASTTPKPEASTTPKPTPTTIKSAFTTPKPEASTTPKPEASTTPKPEASTTPK 499
QY 534 PAPTTPKPEASTTPKPEASTTPKPTPTTIKSAFTTPKPEASTTPKPEASTTPKPEASTTPKPEASTTPK 593
DB 500 PAPTTPKPEASTTPKPEASTTPKPTPTTIKSAFTTPKPEASTTPKPEASTTPKPEASTTPKPEASTTPK 559
QY 594 TPBELAPTTPPEPTPTTPPEPAPTTPKAAAPTTPKPEASTTPKPEASTTPKPEASTTPKPEASTTPK 653
DB 560 TPBELAPTTPPEPTPTTPPEPAPTTPKAAAPTTPKPEASTTPKPEASTTPKPEASTTPKPEASTTPK 619
QY 654 TAPTTPKGTAPTLKPEPAPTTPKPKAPKELAPTTTKEPTSTTSDKPAFTTPKGTAPTLKPEPAPTTPK 713
DB 620 TAPTTPKGTAPTLKPEPAPTTPKPKAPKELAPTTTKEPTSTTSDKPAFTTPKGTAPTLKPEPAPTTPK 679
QY 714 EPAPTTPKPEASTTPKPEASTTPKPTPTTIKSAFTTPKPEASTTPKPEASTTPKPEASTTPKPEASTTPK 773
DB 680 EPAPTTPKPEASTTPKPEASTTPKPTPTTIKSAFTTPKPEASTTPKPEASTTPKPEASTTPKPEASTTPK 739
QY 774 KETAPTTPKPEASTTPKPEASTTPKPTPTTIKSAFTTPKPEASTTPKPEASTTPKPEASTTPKPEASTTPK 833
DB 740 KETAPTTPKPEASTTPKPEASTTPKPTPTTIKSAFTTPKPEASTTPKPEASTTPKPEASTTPKPEASTTPK 799
QY 834 PTPKALENSPKPEGVPTTKTAAATKPEMTTAKDKTTERDLRTTPEITTAAPKMTKETAT 893
DB 800 PTPKALENSPKPEGVPTTKTAAATKPEMTTAKDKTTERDLRTTPEITTAAPKMTKETAT 859
QY 894 TTEKTTESKITATTTQVSTTTQDTPPKIITTLKTIILAPKVTTKKTIITTEIMNKPEE 953
DB 860 TTEKTTESKITATTTQVSTTTQDTPPKIITTLKTIILAPKVTTKKTIITTEIMNKPEE 919
QY 954 TAKPKDRATNSKATTPKPKPTKAPKPKTSKKPKTWPVRVKPTTTPPKMTSTMPELN 1013

Db	920	TAKPKDRATNSKATTPKQKPTKAPKKTSTKKPTMPRVKPKTTPRKMSTMPELN	979
Qy	1014	PTSRIAEAMLQTTTRPNQTPSKLYEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMD	1073
Db	980	PTSRIAEAMLQTTTRPNQTPSKLYEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMD	1039
Qy	1074	YLPRVFNQGIINPMLS	1090
Db	1040	YLPRVFNQGIINPMLS	1056

Search completed: October 13, 2004, 11:52:45
Job time : 117.895 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 119.845 Seconds
(without alignments)
5233.063 Million cell updates/sec

Title: SEQ1-B
Perfect score: 5826
Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....DMDYLPRVFNQGIINPMLS 1090

Scoring table: BLOSUM62
Gapop 10.0., Gapext 0.1

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: -1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 02: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5811	99.7	1404	Q92954	Q92954 homo sapien
2	5788	99.3	1404	Q9BX49	Q9BX49 homo sapien
3	3128.9	53.7	933	Q62MZ5	Q62MZ5 homo sapien
4	3128.9	53.7	933	BAD18580	BAD18580 h cdna fl
5	2700.5	46.4	1054	Q9JN99	Q9JN99 mus musculus
6	1394.5	23.9	5179	1 MUC2_HUMAN	Q02817 homo sapien
7	1321.9	22.7	1225	Q9VR49	Q9VR49 drosophila
8	1314.1	22.6	1761	Q7KTF6	Q7KTF6 drosophila
9	1314.1	22.6	1761	AA564673	AA564673 drosophila
10	1243.6	21.3	3150	Q7PMD5	Q7PMD5 anopheles g
11	1187.5	20.4	3409	Q6SSE6	Q6SSE6 chlamydomon
12	1187.5	20.4	3409	AA507044	AA507044 chlamydom
13	1127.2	19.3	1664	1 SLPI_CLOTM	Q06852 clostridium
14	1113.7	19.1	1349	Q8WWQ4	Q8WWQ4 homo sapien
15	1103.7	18.9	1795	Q76S94	Q76S94 drosophila
16	1098.1	18.8	3889	Q6SSE8	Q6SSE8 chlamydomon
17	1098.1	18.8	3889	AA507042	AA507042 chlamydom
18	1090.2	18.7	3432	Q8IR51	Q8IR51 drosophila
19	1090.2	18.7	3432	Q8IR52	Q8IR52 drosophila
20	1059.8	18.2	1079	Q9N4S7	Q9N4S7 caenorhabdi
21	1035.3	17.8	9234	Q7KTF5	Q7KTF5 drosophila
22	1035.3	17.8	9234	AA10531	AA10531 drosophila
23	1031	17.7	2284	Q9VPG1	Q9VPG1 drosophila
24	1028.7	17.7	1607	Q8HQ5	Q8HQ5 phytophor
25	1028.7	17.7	1607	AA74661	AA74661 phytophor
26	1027	17.6	5703	1 MUSE_HUMAN	Q9NC84 homo sapien
27	1022.7	17.6	972	Q7QKK7	Q7QKK7 anopheles g
28	1019.8	17.5	23015	Q8IQ18	Q8IQ18 drosophila
29	1019.8	17.5	23015	AA10358	AA10358 drosophila
30	1004.7	17.2	1489	Q96449	Q96449 phytophor
31	1002.8	17.2	1274	Q20007	Q20007 caenorhabdi

RESULT 1

Q92954 ID Q92954 PRELIMINARY; PRT; 1404 AA.

AC Q92954; 01-FEB-1997 (Tremblrel. 02, Created)

DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)

DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)

DE Megakaryocyte stimulating factor.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C., Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J., Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T., Bhatia S., Kriz R., Hewick R., Clark S.C.;

RA "Purification, Biochemical Characterization, and Cloning of a Novel Megakaryocyte Stimulating Factor that has Megakaryocyte Colony Stimulating Activity";

RL Blood 78:279-279(1991).

RN [2]

RP SEQUENCE FROM N.A.

RA Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P., Fitzgerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R., Jacobs K., Turner K.;

RA "A Comparison of Vitronectin and Megakaryocyte Stimulating Factor";

RL (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J., Mosher D.F. (eds.);

RL BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS., pp.45-52, Elsevier Science Publishers B.V. (1993).

RN [3]

RP SEQUENCE FROM N.A.

RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C., Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J., Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T., Bhatia S., Kriz R., Hewick R., Clark S.C.;

RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; U70136; AA09089.1; ;

DR HSP; P04004; 1000.

DR Genew; HGNC:9364; PRG4.

DR GO; GO:0008283; P:cell proliferation; TAS.

DR InterPro; IPR000585; Hemopexin.

DR InterPro; IPR001212; Somatomedin_B.

DR Pfam; PF00045; Hemopexin; 2.

DR Pfam; PF01033; Somatomedin B; 2.

DR PRINTS; PR00022; SOMATOMEDINB.

DR SMART; SM00120; HX; 2.

DR SMART; SM00201; SO; 2.

DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.

DR PROSITE; PS00524; SOMATOMEDIN_B; 2.

DR PROSITE; PS00524; SOMATOMEDIN_B; 2.

DR SEQUENCE 1404AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

32 993.1 17.0 2112 2 Q9VEL9

33 992.9 17.0 34350 2 Q8WZ42

34 992.5 17.0 7962 2 Q10465

35 992.3 17.0 10578 2 Q8ISF5

36 990.3 17.0 18519 2 Q8ISF6

37 990.3 17.0 18519 2 Q8ISF7

38 974.5 16.7 2187 2 P70670

39 964.4 16.6 926 2 Q9VVG2

40 964.1 16.5 1458 2 Q757N5

41 964.1 16.5 1458 2 AAS2662

42 960.4 16.5 1720 2 A8I486

43 957.2 16.4 2042 2 Q767L8

44 957.2 16.4 2042 2 BAD08434

45 956.8 16.4 1121 2 Q7Z884

Q9vel9 drosophila

Q8wz42 homo sapien

Q10465 homo sapien

Q8isf5 caenorhabdi

Q8isf6 caenorhabdi

Q8isf7 caenorhabdi

P70670 mus musculus

Q9vvg2 drosophila

Q757n5 ashbya goss

AAS2662 ashbya go

A8I486 plasmodium

Q767L8 sus scrofa

BAD08434 sus scrofa

Q7Z884 candida alb


```

Db 430 TKTPAAKEMTTAKDKTTERDLATTPPTTAAAPKMKETAATTEKTESKITATTQV 489
QY 911 TSTTTQDTPFKITLTKTTLAPKVTTKKTTTTEIMNKPEETAKPKDRATNSKATTPK 970
Db 490 TSTTTQDTPFKITLTKTTLAPKVTTKKTTTTEIMNKPEETAKPKDRATNSKATTPK 549
QY 971 POKPTKAPKPTSTKPKMPVRKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTTPN 1030
Db 550 POKPTKAPKPTSTKPKMPVRKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTTPN 609
QY 1031 QTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFPVETPDMDYLPRVFNQGIINPMLS 1090
Db 610 QTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFPVETPDMDYLPRVFNQGIINPMLS 669

RESULT 4
BADI8580 PRELIMINARY; PRT; 933 AA.
AC BADI8580;
DT 12-MAY-2004 (TREMBlrel. 27, Created)
DT 12-MAY-2004 (TREMBlrel. 27, Last sequence update)
DT 12-MAY-2004 (TREMBlrel. 27, Last annotation update)
DE CDNA FL16561 fis, clone SINOVA003981, moderately similar to Homo
DE sapiens proteoglycan 4, (megakaryocyte stimulating factor, articular
DE superficial zone protein, camptodactyly, arthropathy, coxa vara,
DE pericarditis syndrome) (PRG4).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Synovial membrane tissue;
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuna M., Kawakami B., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isogai T.;
RA "NEDO human cDNA sequencing project.";
RT Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL, AK131434; BADI8580.1;
SQ SEQUENCE 933 AA; 102512 MW; B883773C5BC21A23 CRC64;

Query Match 53.7%; Score 3128.9; DB 2; Length 933;
Best Local Similarity 54.2%; Pred. No. 1.1e-53;
Matches 618; Conservative 0; Mismatches 1; Indels 521; Gaps 3;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSACRGEGYSRATNCNCHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
QY 61 KRVTCTAELSCRGCFESPERGECDCDAQCKYDKCCPDFESFCA----- 105
Db 26 -----ELSCRGCFESPERGECDCDAQCKYDKCCPDFESFCAEVHNPTSPSSKAP 79
QY 106 -----EHSVSENQESSSSSSSSSSSSSTI 130
Db 80 PPGASQTIKSTTKRSPKPPNKKXKTKVIESEEITEHSVSENQESSSSSSSSSSSTIR 139
QY 131 KIKSKNSAANRELQKLKVDNKNKTKKPTKPPVWDVAGSLGDLGDFKVTTPDTST 190
Db 140 KIKSKNSAANRELQKLKVDNKNKTKKPTKPPVWDVAGSLGDLGDFKVTTPDTST 199
QY 191 TQHNKVTSPKITTAKPINRPSLPNPSDTSKTSLSLTVNKEVTKETTTNKQISTDG 250
Db 200 TQHNKVTSPKITTAKPINRPSLPNPSDTSKTSLSLTVNKEVTKETTTNKQISTDG 259
QY 251 KEKTTSAKETOSIEKTSKADLAPTSKVLAKTPKPAETTTKGPALTTPKEPTTPKPEAS 310
Db 260 KEKTTSAKETOSIEKTSKADLAPTSKVLAKTPKPAETTTKGPALTTPKEPTTPKPEAS 319

```

```

QY 311 TTKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEP 370
Db 320 TTKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKE 370
QY 371 APTTTKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 430
Db 371 ----- 370
QY 431 EPAPTAPKAPAPTTTPKEPAPTTTPKEPAPTTTKPSPTTPKEPAPTTTKSAPTTTPKEP 490
Db 371 ----- 370
QY 491 TTKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 550
Db 371 ----- 370
QY 551 APTAPKEPAPTTTPKETAPTTPPKKLTTPTEKLAPTTTPKEPAPTTTPKEPAPTTTPKEP 610
Db 371 ----- 370
QY 611 PEPAPTTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGTAPTTLKEP 670
Db 371 ----- 370
QY 671 APTTPKAPAPKELAPTTTPKETPTSTTSDKAPATTTPKGTAPTTPKEPAPTTTPKEPAPTTPKG 730
Db 371 ----- 370
QY 731 TAPTTPKEPAPTTTPKAPKAPKELAPTTTKGPTSTTSDKAPATTTPKETAPTTPKEPAPTTPK 790
Db 371 ----- 370
QY 791 KPAPTTPETPPPTTSVSTPTTTTKBPTTIHKSPDSTPSELSAETPKALENSPKPGVPT 850
Db 371 -FAPTTPEPTPTTSVSTPTTTTKBPTTIHKSPDSTPSELSAETPKALENSPKPGVPT 429
QY 851 TKTPAAKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTTKTESKITATTQV 910
Db 430 TKTPAAKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTTKTESKITATTQV 489
QY 911 TSTTTQDTPFKITLTKTTLAPKVTTKKTTTTEIMNKPEETAKPKDRATNSKATTPK 970
Db 490 TSTTTQDTPFKITLTKTTLAPKVTTKKTTTTEIMNKPEETAKPKDRATNSKATTPK 549
QY 971 POKPTKAPKPTSTKPKMPVRKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTTPN 1030
Db 550 POKPTKAPKPTSTKPKMPVRKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTTPN 609
QY 1031 QTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFPVETPDMDYLPRVFNQGIINPMLS 1090
Db 610 QTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFPVETPDMDYLPRVFNQGIINPMLS 669

RESULT 5
Q9JW99 PRELIMINARY; PRT; 1054 AA.
ID Q9JW99
AC Q9JW99;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Mus musculus.
GN Name=Prc4;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20573856; PubMed=11124536;
RA Ikegawa S., Sano M., Koshizuka Y., Nakamura Y.;
RT "Isolation, characterization and mapping of the mouse and human PRG4
RT (proteoglycan 4) genes.";

```


CC bronchus, cervix and gall bladder.
 CC -!- PTM: All cysteine residues are involved in intrachain or
 CC interchain disulfide bonds (By similarity).
 CC -!- POLYMORPHISM: The number of repeats is highly polymorphic and
 CC varies among different alleles.
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
 CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
 CC -!- SIMILARITY: Contains 2 WFCC domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; L21998; AAB95295.1; -
 CC EMBL; M74027; AAB59875.1; -
 CC EMBL; M94131; AAB59163.1; -
 CC EMBL; M94132; AAB59164.1; -
 CC PIR; A43963; A43932. -
 CC Genew; HGNC:7512; MUC2.
 CC MIM; 158370. -
 CC InterPro; IPR002919; Cys-rich_TIL.
 CC InterPro; IPR006208; Cys_knot.
 CC InterPro; IPR006207; Cys_knot_C.
 CC InterPro; IPR006209; EGF-like.
 CC InterPro; IPR001007; WFC C.
 CC InterPro; IPR001846; WFC D.
 CC Pfam; PF00007; Cys_knot; 1.
 CC Pfam; PF01826; TIL; 1.
 CC Pfam; PF00093; VMC; 1.
 CC Pfam; PF00094; VMD; 4.
 CC SMART; SM00041; CT; 1.
 CC SMART; SM00214; VMC; 2.
 CC SMART; SM00216; VMD; 4.
 CC PROSITE; PS01185; CTCK_1; 1.
 CC PROSITE; PS01225; EGF_1; 1.
 CC PROSITE; PS00222; EGF_1; UNKNOWN_1.
 CC PROSITE; PS01208; WFC_1; 2.
 CC PROSITE; PS0184; WFC_2; 2.
 CC Glycoprotein; Repeat; Signal.
 KW SIGNAL 1 20
 FT CHAIN 21 5179
 FT DOMAIN 1401 1747
 FT REPEAT 1401 1416
 FT REPEAT 1417 1432
 FT REPEAT 1433 1448
 FT REPEAT 1449 1464
 FT REPEAT 1465 1471
 FT REPEAT 1472 1478
 FT REPEAT 1479 1494
 FT REPEAT 1495 1517
 FT REPEAT 1518 1533
 FT REPEAT 1534 1556
 FT REPEAT 1557 1572
 FT REPEAT 1573 1596
 FT REPEAT 1597 1612
 FT REPEAT 1613 1635
 FT REPEAT 1636 1651
 FT REPEAT 1652 1675
 FT REPEAT 1676 1683
 FT REPEAT 1684 1699
 FT REPEAT 1700 1715
 FT REPEAT 1716 1731
 FT REPEAT 1732 1747
 FT DOMAIN 4815 4886
 FT DOMAIN 4924 4991
 FT DOMAIN 5075 5160
 FT DISULFID 5075 5122
 FT DISULFID 5089 5136
 FT DISULFID 5098 5152

FT DISULFID 5102 5154 By similarity.
 FT DISULFID ? 5159 By similarity.
 FT CARBOHYD 163 163 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 423 423 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 670 670 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 770 770 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 894 894 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1139 1139 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1154 1154 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1215 1215 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1230 1230 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1245 1245 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1787 1787 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 1820 1820 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4339 4339 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4351 4351 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4362 4362 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4373 4373 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4422 4422 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4438 4438 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4502 4502 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4616 4616 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4627 4627 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4752 4752 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4787 4787 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4881 4881 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4888 4888 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4955 4955 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 4970 4970 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 5019 5019 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 5038 5038 N-linked (GlcNAc. . .) (Potential).
 FT CARBOHYD 5069 5069 N-linked (GlcNAc. . .) (Potential).
 FT CONFLICT 1351 1351 H -> L (in Ref. 3).
 FT CONFLICT 1412 1412 T -> S (in Ref. 3).
 FT CONFLICT 1449 1449 L -> P (in Ref. 3).
 FT CONFLICT 1504 1504 M -> T (in Ref. 3).
 FT CONFLICT 4192 4192 G -> S (in Ref. 2).
 SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;
 Query Match 23.9%; Score 1394.5; DB 1; Length 5179;
 Best Local Similarity 20.2%; Pred. No. 6, 4e-19;
 Matches 449; Conservative 88; Mismatches 495; Indels 1195; Gaps 74;
 QY 6 LPIYLLLLSVFVIQVSSQLSSCAGRCG-----EG-----YSR 40
 DB 502 LQVQLAPVQLFVTLDAQSQ--GQVQGLCGNFNLEGGDDFKTAGLVEATGAGFANTWKA 559
 QY 41 DATCN-----CDY--- 48
 DB 560 QSTCHDKLWLDPPCLNTESSANYAHWCSSLLKKTETPFGRCHSAVDPAEYKRYKDYTC 619
 QY 49 NCOHYMEC-----CPDFKRYCTA-----ELSK 71
 DB 620 NQONNEDCLCAALSSVARACTAKGVMLGWRHVCKNDVGCPCNSQVFLYLLTTQQTCTCR 679
 QY 72 -----GRCFESP-----ERGR-----ECDC-----DAQCKYDK 95
 DB 680 SLSEADSHCLGFPAPVDGCGCPDHTFLDEKGCRCVPLAKCSCYHRGLYLEAGVWVQRER 739
 QY 96 C----- 96
 DB 740 CVCRDGRHLRCQIRLIGQSTAPKIHMDCSNLATATSKPRALSCQTLAAGYYHTCEVSG 799
 QY 97 --CPD----- 99
 DB 800 CVCPCDGLMDDRGGCVVEKECPCHNNNDLYSSGAKIKVDCNTCTCKGRWVQTQVCHGT 859
 QY 100 -----YESFCAEEHS-----VSENQSSSSSSSSSSSS 126
 DB 860 CSYIGSGHYITDGYKYPDFDGHCSYVAVQDYCGQSSSLGFSFIITENPCGTTGTCCKA 919
 QY 127 STIWKTKSSKNSAANRELQKKLVKDNKK----- 155

Db 920 IKIFMGRT-----ELKLEDKHRAVIOQDEGHVAVYTTREVGOYLVSSTGII 967
QY 156 ---NRT---KKKPTPKPPVDEAGSL-----DNGDFKV-----183
Db 968 VIWDKETTIVFIKLAPSYKGTV-----CGLCGNFDRSNNDFTRDHMVVSSSELDGNSWK 1022
QY 184 ---TTPDITSTTQ-----HNKYSTSP-----200
Db 1023 EAPTCDDVSTNPEPCLNPHRSWAQKQCSILKSSVFSICHSKDPKPFYACVHDSCSC 1082
QY 201 ---TAKPIN-----KI 202
Db 1083 DTGGDCEFCFSAVASAQBCTKEGACVFWRTPOLCPIFCDYNNPPECWHVEPCGNRSF 1142
QY 203 ---TAKPIN-----PRPSL 214
Db 1143 ETCRTINGIHSNISVYLEGCVPRCPKPRIVEEDLKKCVADKCGCVYVEDTHYPGASV 1202
QY 215 P-----PNSDTSKE-----TSLTVNK 230
Db 1203 PTEETCKSCVCTNSQVWCPRPEGKILNOTQDGFYWEICGPNGTVEKHFNICGII 1262
QY 231 ET-----TVETKETITNKQTSIDGKEKITSK-----258
Db 1263 STLTFTTITLPTTPTSTFTTTTTTTTTTSSVLSLTPKLCCLMSDWINEDHPSSGSDGD 1322
QY 259 ---ETOSIEKTSKD-----270
Db 1323 REFFDGVCGAPDIECRSVKDHLSLEQHGQKQVQCDVSVGFICKNEDQFNGPGLCYDY 1382
QY 271 ---LAPTSKVLAKPTKABETTKGPAITTPKEPTTTTPKESPASTTPKEPT 317
Db 1383 KIRVNCWPMDCIITPSPPTTTPSPPTTTTTLPTTTPSPPTTTTTPPTTTPSPPI 1442
QY 318 PTTIKSAPPTPK-----PAPTITKSAPPTPKAPPTTTPKEAPPTPKAPPTTKE 369
Db 1443 TTTTTLPLTTPSPISITTTTPPTTTPSPPTTTPSP-PTTTPSPPTTTTTPPTTTPS 1501
QY 370 P-----APTITKSAP-----TTPKEPA-----PTTP 390
Db 1502 PPMWTPITPPASTTTLPTTTPSPPTTTTTPPTTTPSPPTTTPPTTTPSTTLPTT 1561
QY 391 KKPAPTTPKEPA-TPKEPTTTTKEPAPTTTPKEPAPTTTPKEPAATAPKKAPITPKPA 449
Db 1562 PSPPPTTTTTPPTTTPSPPTTTTTPSPPTTTTTPSPPTTTTTPPTTTPSPPTTTPSP 1621
QY 450 PTTTPKEP-----APTITKPSPTTPKEPAPTTTPKSAPT-----TTPKEAPTITK 493
Db 1622 TTTTPTTPTTTLPTTTPSPPTTTTTPPTTTPSPPTTTPSPPTTTPSPPTTTPPTTTP 1681
QY 494 SAP-----TTPKESPTTTPKEPAPTTTPKEPAPTTTPKAPPTTPKEPAPTTTPKEP-- 542
Db 1682 SSPITTTTPSPPTTTTTPSPPTTTTTPSPPTTTTTPSPPTTTTTPSPPTTTPSPPTTTP 1741
QY 543 -----APTITKAPAPTAPKEAPTTP-----KETAPTTPKKLT-----576
Db 1742 TTMWTLPTTTPSPITPLPSITPTTPSPPTTTPPTP-CVPLCWTGWLSDGKNF 1799
QY 577 -----576
Db 1800 HKPGGDTLIGDVGFGWAANISCRATWVDPVIGQLGQTVVCDVSVGLICKNEDQKSPG 1859
QY 577 -----TTPKEAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP-----608
Db 1860 VIPMAFCLNLYNEINVOCCBCVQPTTMTTITENPTPTTPTTPTTPTTPTTPTTPTTPT 1919
QY 609 -----TTPBEAPT-----TPKAAAPTNP-----KE 629
Db 1920 TTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 1979
QY 630 PAPTTPKEP-----APTTPKEAPT-TPKEAPTTPKGTATPTLKEPAPTTPKAPAPKEL 683
Db 1980 PPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2039

QY 684 APTTTPKEP-----TSTTSKDPAPT-----TPKG 706
Db 2040 TTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2099
QY 707 T-----APTTPKEP-----TTPKEAPT-----726
Db 2100 TQPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2159
QY 727 ---TPKGT-APT-----LKEPAPTTPKPKAPKELAPTTPKGP-----760
Db 2160 VTPPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2219
QY 761 ---TSTTSKDPAP-----TTPKETAPTTPKEPAP-----TTPKAPAP 794
Db 2220 TPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2279
QY 795 T-----TPETPP-----PTTSEVS-----TPTTTPKEPTTTHKS 822
Db 2280 TPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2336
QY 823 PDESTPELSAEPKPALENSKEP-----GVPT-TKTAAATKPEMTTAKOKTTE 871
Db 2337 PTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2396
QY 872 RDLR---TTP-ETTAAAPKMT---KETATTETKTESKI-----TATTTOVST 913
Db 2397 TGTQPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2456
QY 914 T-----TQDTPPEKIT-----LKTTLAPKVVTT-KKT 941
Db 2457 TTVPTPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2516
QY 942 ITTTEIMNKPEETAKPKDRATNSKATTP-----KPKETKAP-----978
Db 2517 PTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2576
QY 979 -KKTSTKPKTMRVRKPKTTPTRKMTSTMPELNPTSRABAMLQTTT-----1027
Db 2577 TPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2628
QY 1028 --RPNQTPNSKLVEVNPKSEDAG-----GAEGTPEMLRAPHV 1063
Db 2629 TQPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2688
QY 1064 FMPEVTP 1070
Db 2689 VTPPTPT 2695

RESULT 7
QYVR49 PRELIMINARY; PRT; 1225 AA.
AC QYVR49;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE CG3047-PA.
GN Name=Sgs1; ORFNames=CG3047;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20136006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanosides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards J.R., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vande W.M., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.H., Blazek J.R., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,

RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler C., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöckel A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jastali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattlei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merklov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,
RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*
RT melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.,
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
RT a genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084 (2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Mira S., Crosby M.A., Muncall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochownik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Battencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.,
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).
RN [5]
RP SEQUENCE FROM N.A.
RG FLYBASE;
RN Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RG FLYBASE;

RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF003575; AAF50957.3; --
DR FLYBASE: FBgn003372; Sgsl
SQ SEQUENCE 1225 AA; 127735 MW; A99AF9D4404C79F4 CRC64;
Query Match 22.7%; Score 1321.9; DB 2; Length 1225;
Best Local Similarity 26.8%; Pred. No. 2.7e-18;
Matches 358; Conservative 118; Mismatches 419; Indels 441; Gaps 51;
QY 10 LLLLLLVFIQV-----SSODLSSCAGCGEGYSDATCNDYNOCHMECCPDF 60
DB 6 IFLTIVLLIIVKVNKANYVDWDSMD-----GFSHIIPOCGGDIY-----PD- 50
QY 61 KRVCTAELSCKRCFESFERGECDCDAQCKYDKCCPDYEFCAEHSVSENOESSSS 120
DB 51 -----PVQPCDTD----- 58
QY 121 SSSSSSTIWKIKSSKNMAARELOKLLK-----VKDNKKNTKKK-TPK 165
DB 59 ---SNPTTKPKQTKRPKSTRRTTKRPRKTKTKTKTKTKTKTKTKTKTKTKTK 115
QY 166 PPVVDKAGSLDNGDFKVTTPDT-----STQHNVKVS-----TSPKITTAKP 207
DB 116 TP-----DTDSPIITGAECTCSDRITTSSTDSITDRTTNTDWTPLCLDTPP 166
QY 208 INRPSPNPNSDTSKETSITVKNKETTVE-----TKETTTNKOTSDGKKT--T 255
DB 167 CT-----CSBESSTAIPESSPCIDTIVTSPCTQETTTPTCTCGTQTTPCT 216
QY 256 SAKETQSIEKTSKDLAPTSKVLAKTPKAETT--TKGPALTTPKEPTTPKPEASTTP 313
DB 217 CQQT-----TTPRSITTTTSRPTTTTPRSTTTTTSRPTTTT 256
QY 314 KEPTPTTIKSAPTPKPEAPTTTKSAP--TPKPEAPTTTKEPAPTTTKEPAPTTTKEPA 371
DB 257 RSTTTTTRRPTTTTTPRSTTTTSCAPTITTPRSTTTTTPRSTTTTTPRSTTTTSCPT 316
QY 372 PTTTKSAPTT-PKEAPPTTP-----KKPAPTTTPE-----PAPTTTPE--- 408
DB 317 RTPRSTTTTTSRPTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTT 376
QY 409 -----PTPTTPKEPAPT-TKEPAPTTTTPKPEAPTPAKKPAITPKE-----PAPTTPK 454
DB 377 TCTSGPTTTTTPRSTTTTSGPTTTTTPRSTTTTSGPTTTTTPRSTTTTSGPTTTTPR 436
QY 455 EPAPTTTKEPSTTPKPEAPTTTKSAPTTTKEPAPTTTKSAP--TTPKEPSTTTTKEPAP 512
DB 437 STTTTSGPTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTTTSGPTT 496
QY 513 TTPKE-----PAPTTTPE-----PAPTTTPE-----PAPTTTKEPAPTTTK 548
DB 497 TTPRSTTTTSGPTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTTT 556
QY 549 KPAPTPAKE-----PAPTTPKETAPTTPKLPPTPEKLPAPTTPKPAITTEELAP 600
DB 557 GPTTTTTPRSTTTTSGPTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTT 616
QY 601 TTPKEPTTTPPE-----PAPTTPKAAANTPKPEAPTTTPE-----PAPTTPK 644
DB 617 TSTSGPTTTTTPRSTTTTSGPTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTTT 676
QY 645 E-----PAPTTPKETAP-----TTPKGTAPTLPKEPAPTTPK----- 677
DB 677 STTTTSGPTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTT 736
QY 678 PAPKELAPTTTKEPTSTT-----SDKPAPTTPKGTAPITPKPAITTPKE----- 722
DB 737 TTPRSTTTTSGPTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTTT 796
QY 723 -PAPTTPKGTAPTLPKEPAPTTPK-----PAPKELAPTTTKEPTSTT----- 764
DB 797 GPTTTTTPRSTTTTSGPTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTTTTPRSTTT 856

QY 203 TTAKPINDPSPGPNDSKST-SLVNKKETVETKETTINK-QTSDGKEKITSAKET 260
 Db 150 TILK-:::--TTIKSAETAKKSTHNPTTKKSTLRTTEPTTRKSKSTAKTTREPTTKRETT 205
 QY 261 Q-----SIEKTSADKLAPTSKVIAKPPKAPKATTKGPAITTPKEPT-----P 302
 Db 206 ERTTQEPSTKTTHE--TTAEPATKKTTHPTTQKSTTLRTEPTTRKSKSTAKTTREP 263
 QY 303 TTPKEPASTTPKEPT-----PTTKSA-----PTTKK----- 329
 Db 264 TTKRETTERTKEPSTKTTTHETAEPAKKTTHPTTQKSTTLRTEPTTRKSKSTAR 323
 QY 330 -----EPA-----PTTKSA-----PTTKK----- 344
 Db 324 TTRPTTKRETTERTTQEPSTKTTTHETAEPAKKTTHPTTQKSTTLRTEPTTRK 383
 QY 345 EPAPTTKEPAPTTPKEPAPTTPKEPA-----PTTKSA-----PTTKK----- 371
 Db 384 SSTAKTRE--PTTKRETTERTTKEPSTKTTTHETAEPAKKTTHPTTQKSTTLRIT 441
 QY 372 --PTTKSA-----PTTKKEPAPTTPKEPA-----PTTKKEP 401
 Db 442 BEPTTRKSTARTTREPPTTKRETTERTTQEPSTKTTTHETAEPAKKTTHPTTQKST 501
 QY 402 A-----PTTKKEPAPTTPKEPA-----PTTKKEPAPTTPKEPA--PT 443
 Db 502 TLRITEPTTRKSKSTAKTTREPPTTKRETTERTTQEPSTKTTTHETAEPAKKTTHPT 561
 QY 444 TPX-----EPA-----PTTKKEPAPTTPKEPS----- 465
 Db 562 TQKSTTLRTEPTTRKSKSTARTTREPPTTKRETTERTTQEPSTKTTTHETAEPAKKT 621
 QY 466 --PTTKKEPA-----PTTKSA-----PT-----TTKEP 487
 Db 622 THEPTTQKSTTLRTEPTTRKSKSTARTTREPPTTKRETTERTTQEPSTKTTTHETAEPA 681
 QY 488 A-----PTTKSA-----PTTKKEPAPTTPKEPA----- 511
 Db 682 ATKTTHEPTTQKSTTLRTEPTTRKSKSTAKTTREPPTTKRETTERTTQEPSTKTTTHK 741
 QY 512 -----PTTK-----EPA-----PTTKKPAAPTTPKEPA--- 535
 Db 742 TTEPTTKKTTHEPTTKKSTTLKPTTEPTTRKSTTKTTRPTTKRKTTERTTKEPTTRK 801
 QY 536 -----PTTKKEPAPTTPKEPA--PTAPK-----EPA-----PTTK 564
 Db 802 TTHKTTBEPTTNK--TTTKKTTHEPTTKKSTTLKPTTEPTTRKSTTKTTRPTTKR 857
 QY 565 ETA-----PTTKKLT-----PTT-PEKLAPTTPE 588
 Db 858 ETTERTTQEPSTKTTTHETAEPAKKTTHPTTQKSTTLRTEPTTRKSKSTAKTTRE 917
 QY 589 KPAPTTPPELAPTTPEPT-----PTTPS--EPA-----PTTKPA 621
 Db 918 --PTTKRETTERTTKEPTTRKTTTHKTTTEPTTKKTTHEPTTKKSTTLKPTTEPTTRKT 974
 QY 622 AAPNTPEPAPTTPKEPAPTTPKEPA-----PTTKETA-----PTTKGTA 663
 Db 975 STTKTTRE--PTTKKTTERTTKEPTTRKTTTHKTTTEPTTKKTTTHPTTKKSTT 1032
 QY 664 -----PTTKKEPAPTTPKKAPELA-----PTTKKEPTSTSKPA-----PTT 703
 Db 1033 LKPTBEPTTRKSTTKTTRPTTKKVTTERTTREPTTSKTTTHETAEPAKKTTHPTT 1092
 QY 704 PKGTA-----PTTKKEPAPTTPKEPA-----PTTKGTAAPTTLKEPA----- 740
 Db 1093 QKSTTLRTEPTTRKSKSTAKTTREPPTTKRETTERTTKEPTTRKTTTHKTTTEPTTKKT 1152
 QY 741 --PTTKK-----KPAPKEAPTTTKGPTSTTSKPAAPTTPKEAPTTPKEPA----- 785
 Db 1153 HBPTTKKSTTLKPTPEE--PTTRK--TSTTKTTRPTTRKKTTRERTTKEPTTRKTTTHKT 1207

QY 786 -----PTTKP-----KP-----AATTPEPTPTTSEVSTPTTTKEP 816
 Db 1208 TBEPTTKTTTKTTHEPTTKKSTTLKPTTEPTTRKSTTKTTRPTTKRVTTERTTREP 1267
 QY 817 ---TTIHKSPDE-----STPELSAEPTRKALENSPKPEGVPTTKTTPAAKPEMTTAK 866
 Db 1268 TTRKTTTHKTTBEPTTKKTTTKTTHEPTTK--STTLAPTEPTTRKSTTKTTRPTTK 1326
 QY 867 DKTTTERDLR-----TTPETTTAAP----- 885
 Db 1327 RVTTERTTREPTRKTTTHKTTTEPTTKKTTTKTTHEPTTKKSTTLKPTTEPTTRKST 1386
 QY 886 -KMTKETAI---TTEKTTESKITATTQVTSSTTODTTPFKITLLKT-----TTLA 932
 Db 1387 TKTTRPTTKRVTTERTTREPTRKTT--TKTTEPTTKKTTTKTTHEPTTKKSTTLK 1444
 QY 933 P-----KVTTTK-----KTIITTEIMKP-----BETAKPKORATNSKATTPK 970
 Db 1445 PTBEPTTRKSTTKTTRPTTKRVTTERTTREPTRKTTTHKTTTEPTTKKTTTKTTTHE 1504
 QY 971 PQ-----KP-----TKAPKXPT----- 982
 Db 1505 PTTKKSTTLKPTTEPTTRKSTTKAETPTTKRVTTERTTREPTRKTTTHKTTTEPTTK 1564
 QY 983 -----STKKPKTMPVRVKPKTTPTRKMTSTMPELNPTSRIAEAMLQTT----- 1026
 Db 1565 KTTTKTTHEPTTKKSTTL-----KTEPTTRKSTTKTTRPTTRKSTTKTADQTT 1620
 QY 1027 -----TPNOTPNSKLVEVNPKSEDAGAGET-----PHM 1057
 Db 1621 KRTAEMSTTNGEPTSVETTTSSNQSNNTTESITTEBQVHHHHHHHHYHKKPADLGESI 1680
 QY 1058 LLRPH----- 1069
 Db 1681 LPLDLPPLPLPLPWPPLPLPEIPPLPLPTALPPLPLPLPEVNLTAISLPLS 1740
 QY 1070 -PDMDYLPRVNP 1080
 Db 1741 LPLPLPLPOLPN 1752
 RESULT 9
 AAS64673
 ID AAS64673 PRELIMINARY; PRT; 1761 AA.
 AC AAS64673;
 DT 01-APR-2004 (TREMBLrel. 27, Created)
 DT 01-APR-2004 (TREMBLrel. 27, Last sequence update)
 DT 01-APR-2004 (TREMBLrel. 27, Last annotation update)
 DE CG33300-PA (Fragment).
 GN CG33300.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Rogers Y.H., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brannon R.C., Wortman J.R., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Beres P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Clawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,


```

Db 1033 LKPTREPTTKSTTKTTREPTTKKVTERTREPTTSKTTTHETTAEPATKKTTHBPTT 1092
QY 704 PKGTA-----PTTPKEPAPTTPKEPA-----PTTPKGTAPTTLKEPA----- 740
Db 1093 OKSTTLRTEBPTTKSTAKTTRPTTKRETTERTKEPTTKTTTHKTTTBEPTTKTT 1152
QY 741 --PTTPK-----KPAPELAPTTKPTSTTSKAPATPKETAFTTPKEPA----- 785
Db 1153 HEPPTTKSTTLKPTBE-----PTTRK--TSSTTKTREPPTTKRTTERTKBPTRKTTTHKT 1207
QY 786 -----PTTPK-----KP-----APTTPPTPTSEVSTPTTKKEP 816
Db 1208 TEEPTTNKTTTKTTHEPTTKSTLTKTBEPTTKRTSTTKTREPPTTKRTTERTREP 1267
QY 817 ---TTIHKSPDE-----STPELSAEPKALNSKPKEGVPPTKTPAATKPEMTTAK 866
Db 1268 TTKTTTHKTTBEPTTKTTTKTTTKEPTTKK--STTLKPTREPPTTKSTTKTREPPTTK 1326
QY 867 DKTERDLR-----TPETTTAAP----- 885
Db 1327 RVTERTTREPPTTKTTTHKTTBEPTTKTTTKTTTKEPTTKSTTLKPTREPPTTKST 1386
QY 886 -KMKETAT--TTEKTESKITATTQVSTTTTODTTPPKITLKT-----TTLA 932
Db 1387 TKTTREPPTTKRVTERTTREPPTTKTT--THKTTBEPTTKTTTKTTTKEPTTKSTTLK 1444
QY 933 P-----KVTTK-----KLTITTEIMNKP-----BETAKPKDRATNSKATTPK 970
Db 1445 PTEPTTKRTSTTKTREPPTTKRVTERTTREPPTTKTTTHKTTBEPTTKTTTKTTTKE 1504
QY 971 PQ-----TKAPKKPT----- 982
Db 1505 PTTKSTTLKTEBPTTKSTTKATREPPTTKRVTERTTREPPTTKTTAKTTEBPTTK 1564
QY 983 -----STKKPKTMPVRPKPTTPPKMTSTWPELNPTSRABAMLOTT----- 1026
Db 1565 KTTTKKTTHEPTTKSTTL-----KTEBPTTKSTTKTREPPTTKSTSVKTTADQTT 1620
QY 1027 -----TRPNQTPNSKLVEVNPKESEDAGGAEGT-----PHM 1057
Db 1621 KETTAEMSTTNOEPTSVETTINSQNSNTTETSTTBEOVHHHHHHHHYHKPADJGPSI 1680
QY 1058 LLRPH-----VFWEVPT 1069
Db 1681 LPLFDLPPLPLPWPPLPLPEPLPLPLPTALPLPLPLPLPLPLPLPLPLPLPLPLPL 1740
QY 1070 -PMDYLPRVEN 1080
Db 1741 LPNLPPLPLQLEN 1752

```

RESULT 10
Q7PMD5

ID Q7PMD5 PRELIMINARY; PRT; 3150 AA.
AC Q7PMD5;
DT 01-MAR-2004 (Tremblrel. 26, Created)
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE ENSANGP0000004655 (Fragment).
GN Name=ENSANGP000000003651;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.

```

DR EMBL; AAAB01008980; EAA13969.2; --
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004872; P:receptor activity; IEA.
DR GO; GO:0008151; P:cell growth and/or maintenance; IEA.
DR InterPro; IPR006770; OGF:recept.
DR Pfam; PF04680; OGF_r_III; 80.
FT NON_TER 1
FT NON_SEQUENCE 3150 3150
SQ SEQUENCE 3150 AA; 322879 MW; 3C7B3DD441C58C839 CRC64;

Query Match 21.3%; Score 1243.6; DB 2; Length 3150;
Best Local Similarity 19.9%; Pred. No. 3e-16;
Matches 402; Conservative 111; Mismatches 397; Indels 1114; Gaps 74;

QY 111 SENQSSSSSSSSSSSSIIWKIKSSKNGAANRELOKLVKDNKKKTKKTPKPPVVD 170
Db 870 TESTDTTMSASTPEPST-----TSGTTR-----TTTPTPTPTDTTMS 907
QY 171 EAGS-----GLDNGDFKVTTPDTSTTQHN----- 194
Db 908 SASPEPSTTPGTTTTPTRPTSTESTDTTMSASTPEPSTTPGTTTTPTRPTSTESTD 967
QY 195 -----KVSTPKITTAKPINP-----RPSLP 215
Db 968 TTMSSASTPEPSTTPGTTTTPTRPTSTESTDTTMSASTPEPSTPKGTTTTPTRPTT 1027
QY 216 PMSDTSKETSILVKNKETTVEKTTTNNKQSTDGKKTSA----- 257
Db 1028 ESTDTMSSASTPEPSTTPGTTTTPTRPTSTESTDTTMSASTPEPSTTPGTTTTPTR 1086
QY 258 --KETQSLEKTSKADLAP---TSKVLAKTPKAST-----TTKGPAL 294
Db 1087 RPTSTESTDTTMSASTPEPSTTPGTTTTPTRPTSTESTDTTMSASTPEPSTTPGTT 1146
QY 295 TTPKSEPT-----TTPKEPAS----- 310
Db 1147 TTPTRPTSTESTDTTMSASTPEPSTTPGTTTTPTRPTSTESTDTTMSASTPEPSTTP 1206
QY 311 -----TTPKEP-----TPTTIKSAPTTP-----KEPA-- 332
Db 1207 GTTTRTTPRPTSTESTDTTMSASTPEPSTTPGTTTTPTRPTSTESTDTTMSASTPE 1266
QY 333 PTTTKSAPTTP-----KEPA 347
Db 1267 PGTTTTPRPTSTESTDTTMSASTPEPSTTPGTTTTPTRPTSTESTDTTMSASTPE 1326
QY 348 PTTTKAPATTTPKEAPT-----TTKEPA--PTTTKSAPTTP----- 382
Db 1327 PTTTPDTTTPRPTSTESTDTTMSASTPEPSTTPGTTTTPTRPTSTESTDTTMSA 1386
QY 383 KEPAP-----TTPKKAPAT-----TPKEPAPTTPKEPTTPTPKEAPT----- 420
Db 1387 STPEPSTTPGTTTTPRPTSTESTDTTMSASTPEPSTTPGTTTTPTRPTSTESTD 1446
QY 421 -----TKEPAP-----APKKAP-----TTPKEPAP--- 450
Db 1447 MSSASTPEPSTTPGTTTTPRPTSTESTDTTMSASTPEPSTTPGTTTTPTRPTSTES 1506
QY 451 -----TTPKEPATTTPKSPPTTPKEPA-----PTTT 477
Db 1507 TDTTMSAYTPDPGTPGTTTTPRPTSTESTDTTMSASTPEPSTTPGTTTTPTRPTST 1565
QY 478 KSAPT-----TKEPA--PTTTKSAPTTP-----KEPSTTTKEPAPTTPKE 517
Db 1566 ESTDTTMSASTPEPSTTPGTTTTPRPTSTESTDTTMSASTPEPSTTPGTTTTPTR 1625
QY 518 PAP-----TTPKKAP-----TTPKAP-----TTPK 532
Db 1626 PRTSTESTDTTMSASTPEPSTTPGTTTTPRPTSTESTDTTMSASTPEPSTTPGTTT 1685
QY 533 EPAPT-----TTPKEPATTTPKPA-----PTAP-----KEAPT 561
Db 1686 RPTPTDSTMSMSESTPEPSTTPGTTTTPRPTSTESTDTTMSASTPEPSTTPGTTT 1745

```


Db 110 ABWPSRCALPAVQGYGYDEYIDATVYSGNSLRPFYSTCRYPASSDDPYAFLEPYSMRWVD 169
 QY 17 -----FVIOQVSS-----QDLSSCAGRCGEGYSRD 41
 Db 170 DFGIYQPVLDGDMWGETVNTVDSNLKWIWDPAWTAQCAWLGGQDF-----RD 217
 QY 42 ATCN-----CDYNCOHYMECCPDFKRVCTAELSCKGRCFESFERGECDCDAQCKKYD-- 94
 Db 218 AVNFGTHYCSW---PFVEC-----SSCENYDIA 243
 QY 95 -----KCCPD-----YESFCAB-----E 107
 Db 244 DPYDPKIANGIVPAVITALDPFNASLYIYLDGFMFPAGSLDNVWYLNLAYNFICGSLP 303
 QY 95 -----KCCPD-----YESFCAB-----E 107
 Db 304 ANLPTLLPSLOHALDHCRAVDVGTASLQYGMQOYFSTGQPYE-YCSDGDVGTGSDGTE 362
 QY 108 HSVS-----ENQ-----ESSSSSSSSSSSSSIW 130
 Db 363 YVIGWIPDEWDGDAVAAAASDSSLPWANLRTVRLSNQALYGPPEGLRSASSISS---W 419
 QY 131 KTKSSKNSAANRELKOKLK-----VKDNKNKTK-----159
 Db 420 RLQ-----GNTELCGLPEFAAPINSLLYLGTLGHTHTWVHDPADHTQGCECIAPPPP 473
 QY 160 -----159
 Db 474 PPSPPRRPRRPLPPSPPPPLPPSPVPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 533
 QY 160 -----KKPTKPPVVVDAGSLONG 179
 Db 534 TPVARCIVQGGICDPSMPSP 581
 QY 180 DPKVTPDP---STCHNKVSTSPKITTAKPINRPSLPPNSDTSKETSITVKNKETTET 236
 Db 582 -FHPPSPDPASSVPPSP 620
 QY 237 KETTTNKQTDGKEKTSKETSQIEXTSADLAPTSKVLAKPTPKAETTTKGPALTT 296
 Db 621 -----PPSPAPPSPAPS-----PAPPS 638
 QY 297 KEPTTTTKEPAST-----TPKEPTTTIKSAPTTKEPAPTTTKEPAPTTTKEPAPT 349
 Db 639 POPSSEVPQPSVPPSP 698
 QY 350 TTKEPAPTTPK-----EPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 396
 Db 699 SPSPSPAPSP 758
 QY 397 -----TPKEPAPTTKEPTTTTKEPAPTTKBPAPTTTKEPAPT-----APKBPAPTTK 447
 Db 759 APPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 818
 QY 448 PAPTTPKEPAPT-----TTKEPSPTTTKEPAPTTT 478
 Db 819 PAPSPSPSPAPPSP 878
 QY 479 SAPTTTKEPAPTTKSAPTTKEPSPSTTTTKEPAPTTTKEPAP-----TTP 523
 Db 879 PAPSPLEPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 937
 QY 524 KKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTKBPATPK-----BPA 559
 Db 938 PSPAPSP 997
 QY 560 PTPPKETAPTTPKLTPTTPEKLAPTTPE--KPAPTTPBELAPT-----T 602
 Db 998 PPSAPSPSPNPPSPVPTTSP--GPPSPSPSPAPPSPPTTSPSPSPSPSPSPSPSPSPSP 1056
 QY 603 PPSPTPTTPEPAPTTTKEPAPTTK-----TPKEPAPTTTKEPAPTTTKEPAPT 652
 Db 1057 PAPPLPPSP 1116

QY 653 ETAPTTKGTAPTTL-----KEPAPTTPK 676
 Db 1117 PPAPSPSPSPAPLPPSPDPSPAPPSPMPPLFTSPSPSPSPSPSPSPSPSPSPSPSP 1176
 QY 677 KPAPKELAPTT-----TKEPTSTTSKAPTTTTPKGTAPTTP 712
 Db 1177 PPSQPLEPSPSPSPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 1235
 QY 713 KEAPATTKEPAPTTKGTAPTTLKEPAPTTTTPKAPKELAPT-TTKGTPTSTTSKAP 771
 Db 1236 PSAPSPSPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 1291
 QY 772 TPKEAPTTPKEPAPTTKAPTTTTPETPTT-TSEVSTPTTTTKEPTTHKSPDESTPEL 830
 Db 1292 SP---APSPSPSPAPPSPSPSPAPPSPSPAPPSPAPPSPAPPSPAPPSPAPP 1348
 QY 831 SAPPTPKALE-----NSPKSPGV---PTTKTP-----854
 Db 1349 SPAPSPFLPPQPTPAMPSPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 1408
 QY 855 -----AATKPE-----MTTTA-----865
 Db 1409 AQALLSIPDAANSVFWVSAGLTSVAPSTPPPELLASFCVCSQLTATAISLVGSSRG 1468
 QY 866 -----866
 Db 1469 NNTGSSSGNYISNGGDAAIQIRPAGNSTDRGSSSGSGSGSGSGSGSGSGSGSG 1528
 QY 866 KDKTTERDL-----RTP-----878
 Db 1529 QDGYQIQLSIGGVYTRTVVDRTPSPVSGNVTLSANRIKQEPSVAGEASLNAKSGQA 1588
 QY 879 -----ETTTAAPKMT-----KETATTTEKTES 901
 Db 1589 MLLTISFSEVPAPDPAASLIVTGALVAEWAAADKMTFYVLAMTLPAELVATAAGSSSS 1648
 QY 902 KIT-----904
 Db 1649 SGTSRSGNGNGTAAAAAAPPAGTGTGRRRLQQOQAAAAAPPPSPASSSSSSLSGAATNQ 1708
 QY 905 -----ATTQVSTTTTQDT 918
 Db 1709 QORHVHLLPATAYADAARNPGRNDLSLSVELTDNAVASPAVGEALATTARVTAATYP-- 1766
 QY 919 TPKITTLKTTTLAPKVT-----936
 Db 1767 -----AVAAATTLVAAAASSSPAQAIRAKGSLLOGSYHIQMLTMSLYLASRGVGEYEA 1821
 QY 937 -----TTKKTITTE-----IMNKPEETAKPK 958
 Db 1822 VEFKAVLVGKGNLGPAAEAMPTEKEVTAABQARQVGGDLWPIGNDLLGSSNTTASGS 1881
 QY 959 DRATNSKATTPKQPK-----TK 976
 Db 1882 SSGSSSSSNRP-PRRPPPPPAAGSTGLLFSNADASPPPLAVATPALPAPLSTTAAAAATA 1940
 QY 977 APKPTSTTKKPTKTPRVKPKKTTTPRKMTST--MPBLN-----PTS 1016
 Db 1941 AP-----PRUPSP---PPPAVGSSTGVLPRRHLMOQMLQPPAAVAAPPPPPA 1985
 QY 1017 RIAEAMLOTTTRPNQNTNSKLVE-----1039
 Db 1986 SSSALVLQPSPPPPPPPSQLLIQOASATYVSDMQDLLVTLVVAAMLITAVAGRLIAAVL 2045
 QY 1040 -----VNPKESDAGGAEGETPHMLLRPHVFMPEVTPDMDYLR-----1077
 Db 2046 YRLVWSPE-----PHPFLA-----FPRLETTIAGLLVALTFVSC 2080
 QY 1078 -----VFNQGIINPM 1088
 Db 2081 MALGPAADWHSRTAAAYCVLTIAVVPYAAFLWLALARAMVP-QFTLVEPM 2132

RESULT 13
SLP1_CLOTM STANDARD; PRT; 1664 AA.
AC Q06852; 1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer
DE protein 1).
GN Name=olpB;
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 10682;
RX MEDLINE=93209931; PubMed=8458832;
RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the
RT cellulosomal scaffolding protein cipA and a protein possibly involved
RT in attachment of the cellulosome to the cell surface.";
RL J. Bacteriol. 175:1891-1899(1993).
CC -! SUBUNIT: Assembled into mono-layered crystalline arrays.
CC -! SUBCELLULAR LOCATION: Cell wall.
CC -! SIMILARITY: Contains 4 S-layer homology (SLH) domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X67506; CAA47841.1; -
DR PIR; T18262; T18262.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001119; SLH.
DR Pfam; PF00395; SLH; 3.
DR PROSITE; PS01072; SLH DOMAIN; 2.
KW Cell wall; Repeat; S-layer; Signal.
FT SIGNAL 1 28
FT CHAIN 29 1664
FT DOMAIN 36 763
FT REPEAT 36 191
FT REPEAT 207 363
FT REPEAT 409 565
FT REPEAT 607 763
FT DOMAIN 771 1377
FT DOMAIN 1378 1449
FT DOMAIN 1453 1494
FT DOMAIN 1495 1565
FT DOMAIN 1566 1625
FT DOMAIN 1626 1646
FT SEQUENCE 1664 AA; 178194 MW; 5F396695BA9FE74B CRC64;
Query Match 19.3%; Score 1127.2; DB 1; Length 1664;
Best Local Similarity 21.4%; Pred. No. 2.5e-14;
Matches 369; Conservative 127; Mismatches 427; Indels 798; Gaps 68;
QY 4 KTLPIYLLLLSVFVIQVQSODLSSCAGRGEGYSRDATCNCYNCQHYMECCP----- 58
DB 6 KVLISILLTLL-----IISTSVNKS-----FAEATPSIEMV 37
QY 59 -----DFKRVCTAELSC-----GRCFSFGRGECDC 86
DB 38 LDKTEVHVGDVITATIKVNIRKLQYQNLKIPDPVLPVDPVDPATGEEFT----- 87
QY 87 DAQCKYDKCCP-----DYESFCAEEH-- 108

Db 88 -----DKSMFVNRVLLTNSKYGPTPVAGNDIKSGIINFATGYNLNTAYKSSGIDEHTG 140
QY 109 -----SVSENQESSSSSSSSSSSTTW-----KIKSS 135
Db 141 IIGIEIGFKVLKQNTSIRFEDTSLMPCGASGTSLFDWDAETITGYEVIQPDLIWAEPL 200
QY 136 KNGAANRELOK-KLVKVD-----NKN-----RTKKPT----- 163
Db 201 KDAVALELDKTKVKGDIITATIKIENKMFAGYQNLKIDPTMLEALELETGCSAIAXR 260
QY 164 -----PKPVDDEAGSLDN----- 178
Db 261 TWPVTGTVLQSDNYGKTTAVANDVGAGIINFAYSNLTKYRETGVAEETGIIGKIGPR 320
QY 179 -----GDFKVTTPDT-----STQHNKV-----STSPK 201
Db 321 VLKAGSTAIRFEDTTAMPAGIECTYMDWYGENIKGYSVQPGIEIYVAEGEERPEEPT 380
QY 202 ITTAKTINRPSLP-----PNS-----DTSK----- 222
Db 381 VPTETVPDPTPTVTERPVPSELPSVIMELDKTKVKVGDIITATIKIENKMFAGYQLN 440
QY 223 -----ETSLVNKETTVETKETTNTKOTSTDGKEKTS----- 256
Db 441 IKYDPTMLEALETGSATKATWPTGTV-----LQSDNYGKTTAVANDVGAGIINFA 495
QY 257 -----AKETQSIEKTSADLAPTSKVLAKPTKAEETT----- 289
Db 496 EAVSNLTKYRETGVAEETGIIGKIGFVLKAGSTAI-----RPEDTAMPAGIEGTYMFD 550
QY 290 -----KGPAITTP-----KEPTPT-TPKEPASTTPKEPTPT----- 319
Db 551 WYGENIKGYSVQPGIEIYVAEGEETPEPTPTVDPTPTVTEEPVPSLPDSYVIMELD 610
QY 320 ----- 319
Db 611 KTKVKEGDVITIRVNNIKNLAGYQIGIKYDPKVLKLEAFNIETGDPIDEGTWAPVGGTIL 670
QY 320 -----TIKSAPTTPK 329
Db 671 KNRDYLPTGVAINNVSKGILNFAAYVYVDDYREEGKGEDTGIIIGNIFVLKAEDETTIR 730
QY 330 -----EPATTTKSAPTTPKPAPTTKEPAPTTKEPAPTT 359
Db 731 FFEESMPSGIDGTWMLDNLNRISGVYVQAPKKAAS-----DEPIPTDTPSDEPTPS 785
QY 360 KEPAP-----TTTKEPAPTTKSAPTTKEPAPTTTPKAPTTTPKEPAPT-----TP 406
Db 786 DEPTPSDEPTPSDEPTPSDEPTPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTP 845
QY 407 K-EPTPT-TPKEPAPT-TPKEPAPTTKEPAPTTAPKKA--PTTPKEPAPT-TPKEPAPT 460
Db 846 SDEPTPEETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPEEPIPTD 905
QY 461 TKEPAPTTPKEPAPTTKSAPTTKEPAPTTTPKAPTTTPKEPAPTTPKEPAPTTPKEPAP 520
Db 906 TSDEPTPSDEPTPS-----DEPTPSDEPTPSDEPTPEEPIPTDTPSDEPTPSDEPTP 962
QY 521 TTPKKA-----PTTPKEPAPT-TPKEPAPTTTPKAPTTAPKAPTTAPKAPTTKE 565
Db 963 SDEPTPSDEPTPSDEPTPSDEPTPEEPIPTDTPSDEPTPSDEPTPSDEPTPSD--- 1019
QY 566 TAPTPKLTTPTPKLAAPTTKEPAPT---TPELAPTTPEEPTPT-TPEEAPTTTPKA 621
Db 1020 -----EPTPSDEPTPSDEPTPSD-EPTPSDEPTPSDEPTPEEPIPTDTPS 1061
QY 622 AAPNTPKSAPTTTPKEA--PTTPKEPAPT-TPKETAPTTTPKAPTTTPKAPTTTPKP 678
Db 1062 DEPTPSDEPTPSDEPTPSDEPTPSDEPTPEEPIPTDTPSDEPTPSDEPTPS--DEP 1119
QY 679 APKELAPTTTKEPT-STTSDKAPTTTPKAPTTTPKEPAPTTPKEPAPTTPKAPTTLK 737
Db 1120 TPSD-EPTPSDEPTPSDEPTPEEPIPTDTPSDEPTPSDEPTPS--DEPTPSD 1172

```
QY 738 EPAPT-TPKKAPKELAPTTTSGPTSTSDKAPT---TPKETAPTTKEPAPT-TPKKP 792
D 1173 EPTSETPEPI---PDTSPDEPTSPDEPTSPDEPTSPD-EPTSPDEPTSETPEP 1226
QY 793 APPTTETPPPTSEVSTPTTKEPTTIHKSPPDESPELSAPPTPKALENSKEGCVPTTK 852
D 1227 IPTDTPSDEPTSPD-EPTSPDEPT---PSDEPTP---SDEPTP---SETPEEP-IPDT 1274
QY 853 TPAATKPEMTTAKDKTTERDLRTTPTTAAPKMTKETATTETKTESKITATTQVTS 912
D 1275 PSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEP 1334
QY 913 TTTQDTTPPKITLTKTLLAPKVTTKTITTEIMNKPEETAKPKDRATNSKATTPKQ 972
D 1335 TPSDEPTSPDEPT---PSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDE 1369
QY 973 -KPTKAPKPTGKPKTMRVRKKTTPPKMTS-----T 1008
D 1370 PTPSPTEPEPT-----PTTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 1416
QY 1009 MPLENTSRIABAMLQTTTRNQTPNSKLVEVNPKSDAGGAGETPHMLRPH---VFM 1065
D 1417 SPTPTPTS-----KPTSTAPTIE-EPTSPDVFGAIGGEHRAYLRLGYDPGSGFR 1464
QY 1066 PE----- 1067
D 1465 PERNITRABAAVIFAKLGADESYGASPSYSLDLADTHAAWAIAIKFATISQGLFKGYPDG 1524
QY 1068 -VTPD-----MDYLPRVNPQI-----IINP 1087
D 1525 TFKPDQNTIRABFATVVLHFLTKVKGQEIEMSKLATIDISNP 1565

RESULT 14
Q8WWQ4
ID Q8WWQ4 PRELIMINARY; PRT; 1349 AA.
AC Q8WWQ4;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Mucin 5 (Fragment).
GN Names=MUC5AC;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1];
RP SEQUENCE FROM N.A.
RX MEDLINE=21426417; PubMed=11535137;
RA Escande F., Aubert J.P., Porchet N., Buisine M.P.;
RT "Human mucin gene MUC5AC: organization of its 5'-region and central
RL repetitive region.";
RL Biochem.J 358:763-772(2001).
DR EMBL; AJ298318; CAC83675.1; -.
FT NON_TER 1
FT NON_TER 1349
SQ SEQUENCE 1349 AA; 135599 MW; 4DC3C1544F1E5EBA CRC64;

Query Match 19.1%; Score 1113.7; DB 2; Length 1349;
Best Local Similarity 23.6%; Pred. No. 3.5e-14;
Matches 347; Conservative 78; Mismatches 371; Indels 673; Gaps 42;

QY 109 SVSNOESSSSSSSSSTIWKSKNSAANRELQKLKVKDNKKNETKKKTPKPPV 168
D 1 TTTSTTAPTSTTSTPTSSSTI----- 20
QY 169 VDEAGSLDNGDFKVTPTDSTTQHNKVTSPKXITAKLINRPP-----SLPPNSDPS 221
D 21 -----TSTPQSTTS-----ASTTSITSGPPTSPVPVTTTSTTAPTSTTS 62
QY 222 KETSLVWKETTVETKETITTNKQSTDGKEK-----TTSKETQSI 263
```

```
Db 63 AATTSTISATTTSTTGAPTTSTTSASTASKTSLGTTPTSPITPTTSTTSPPTTSTTASTA 122
QY 264 EKTSAKOLAP-----TSKVLAKPTPKAET-----TTKGPALTTPKEPPTPTT----- 304
D 123 SKTSGPCTTSPVPTTSTIFA---PRTSTTSASTTSTTPGPG-TTP-SPVPTTSTASVSK 177
QY 305 ----- 304
D 178 TSTSHVISIKTHSQPVTRDCHLRCTWKWFDVDFPSPGPHGDKETNNIIRSGBKICR 237
QY 305 -PKE----- 307
D 238 RPEBITRLQCRASHPEVSIHGLQVVQCSREGLVCRNQDQGGPFKMLNVEVRVLCC 297
QY 308 -----PASITP-----KEPT-----PTTIKSAPTTTKEPAPT- 335
D 298 TPKGCPVTSTPTVAPSTPSPGRATSPQSTSSWOKSRTTLVLTSTTSTPTSTTAPTTS 357
QY 336 TKSAPTTTKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK-----SAPTTTKEPA 386
D 358 TTSAPTTTSTTSTTSTPTQSTSSAPTSSITTSAPTSSITTSAPTSSITTSAPTSSITSS 417
QY 387 PTPPKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP-----TTKEPAPT- 428
D 418 PTTTSTGATTSTTAPTSTTSTPTQSTKSTSAATSTSSGTTSPSPVTTTSTASVSKTS 477
QY 429 ----- 429
D 478 TSHVSVSKTHSQPVTRDCHPRCTWKWFDVDFPSPGPHGDKETNNIIRSGBKICRRP 537
QY 430 KE----- 431
D 538 QEITRLQCRASHPEVSIHGLQVVQCSREGLVCRNQDQGGPFKMLNVEVRVLCCETP 597
QY 432 -----PAPTAPKKEPAPT-----TPKEPAPTTPKE 455
D 598 KGCVPVTSVTAPEPLVGEPPAQIQTSTSWOKSRTTLVLTSTTSTTSTTAPTSTT 657
QY 456 PA---PTTTKEPAPTTPKEPAPTTPKSAPTTTKEPAPTTPKSAPTTPKEPAPTTPKEPA 511
D 658 PASIPSTTS---APTSTTAPTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTISA 715
QY 512 PTTTKEPAPTTPKAPAPTTPKEPAPTTPKEPAPTTPKAPAPTTPKAPAPTTPKAPAPTTP 571
D 716 PTTTSTGATTSTTAPTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTT 774
QY 572 KKLPTTPEKLAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPA 631
D 775 TTSPTQSTSSPTSTTPTQSTSSPTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSSA 834
QY 632 PTTTKEPAPTTPKEPAPTTPKEPAPTTPKGTAPTTPKEPAPTTPK-----K 677
D 835 PTSSTTAPTASTTAPTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTT 894
QY 678 PAKKELAPTTPKEPTSTTS-----DKPAP----- 701
D 895 PSP---VPTTSTASVSKTSHVSVSKTHSQPVTRDCHPRCTWKWFDVDFPSPGPHG 951
QY 702 ----- 701
D 952 DKETNNIIRSGBKICRRPEBITRLQCRASHPEVSIHGLQVVQCSREGLVCRNQDQ 1011
QY 702 -----TTPKG-----TAPTTPK----- 713
D 1012 GPFKMLNVEVRVLCCETPCKGCPVTSTPTVAPSTPSPGRATSPQSTSSWOKSRTTLVTT 1071
QY 714 -----EPAPTTPKEPAPTTPKGTAPTTPKEPAPTTPKAPKAPKELAPTTPKGTST 763
D 1072 STTSTPTQSTTSTAPTSTTIPASTPTSTTAPTSTTAPTSTTAPTSTTAPTSTTAPTST 1128
QY 764 TSDKAPAPTTPKEPAPTTPKEPAPTTPKAPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 821
D 1129 TT--LAPTTSTTAPTSTTNSAPTSTTISATSTTISATSTTISATSTTISATSTTSTTPTQSKT 1186
```

```

QY      822 SPDESTPELSAEPKPKALENSKPEKPGVTTTKTAAKPMETTTAKDKTTERDLRTTPTTT 881
DB      1187 SAATSTTSGSGTTP-----SP-----VPTTSTTSASTTSTTSAPTSTT-----SGPGTT 1232
QY      882 TAAPKMTKATTTTETKTSKTAITTTQVTTTQ--DTTPKTIITLTKTTLAPKVTTK 939
DB      1233 PSEVPSTSTSAATSTTSAPTTRTTSAPTSTTSGPGTTPSPVPTTSTTS-APTSTT-- 1289
QY      940 KUITTEINNKBEETAKPKDRATNSKATPKPKAPKP-TSTKKPKMPPVRKPKPT 998
DB      1290 -----TSGPGTTPSPVPTTSTTSAPTSTT-----TSGPGS 1319
QY      999 TTPPKVKTMPELNPTSRIAPMLQTTT 1027
DB      1320 TSPVPTTSTTSA--PTTSTTSASTT 1346

RESULT 15
O76894 PRELIMINARY; PRT; 1795 AA.
AC O76894;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE CG14796-PA (EG:56G7.1 protein).
GN Name=EG:56G7.1; ORFNames=CG14796;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gaber G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benton P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dalke C., Davenport L.B., Davies P.,
RA de Fabios B., DeCher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hoslin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Krat C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reaert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J., Ye J.,
RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).

```

```

RN RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celnikner S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celnikner S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crossy M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradscky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celnikner S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RX FLYBASE;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX FLYBASE;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RX Cadieu, Dreano, Lelaure, Mottier, Galibert F.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RX Benos P.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN EMBL; AEO03421; AAF45644.1; -
RN EMBL; AL031028; CAA19845.2; -
RN FLYBASE; PEGn0025390; CGI4796.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0008061; F:chitin binding; IEA.
DR GO; GO:0006030; P:chitin metabolism; IEA.
DR InterPro; IPR002557; Chitin_bind_PerA.
DR Pfam; PF01607; CBM 14; 2.
DR SMART; SM00494; CHIBD2; 3.
DR PROSITE; PS50940; CHIT_BIND_II; 2.
SQ SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;
Query Match 18.9%; Score 1103.7; DB 2; Length 1795;
Best Local Similarity 21.2%; Pred. No. 7.8e-14;
Matches 382; Conservative 164; Mismatches 444; Indels 813; Gaps 66;
QY 3 WTLPIYLLILLISVFVIQVSSQDLSSCAGRC-----G 35
DB 6 WST-----LIGLLSV-----LSTATSRVQLPSAYPPVLTGYGQPPYAYPPYG 51
QY 36 EGVSR-----DATCNCDCYNC-----OHYMECC 57
DB 52 YGYQAPPVYQPYDYNVFGVSKYPGGGFFVRFDFNNECSNIGIKHPHQQQYYVCK 111

```

```

QY 58 PPKRVCTAELCK-----GRCFBSFERGR-----BCDDAQ-----CKKYDKC 96
Db 112 PD-----CVIFSKRGLESNASSGRCVQHQRHDPHPPQCOQKEGRFFPHDCKVYRC 167
QY 97 -----CPDYE-----SF-----CAEE-----107
Db 168 DKNRQOWLFACPAGTIFSPVERKCLPGQOCFSTISDSGSVIPONCELKPECEAEGTF 227
QY 108 -----HVSBNQESS-----SSSSSSSSTIWK-----131
Db 228 RSPDICALYVTCRLQBSGYLQTRFKPCGNSFDLERKLCRPRSEVDCPFDVFGPVQVPY 287
QY 132 -----TKSSK-----136
Db 288 APQYVPPYPAAPLYEEDDYDTGAREQPALKSEKLQVAAGFEKPSLNVVVLQTTILE 347
QY 137 -----NSAANRELQK-----KLKVKON-----153
Db 348 PSTAYHKYPAYSPSYEYSSHRRGKERAENLEKEGVPKRLKLSINIVQETPATA 407
QY 154 -----KKNRTKKPKTPKPPVVDGAGSLDNGDFKVTTPDTSST 191
Db 408 ATTREPLNDINKYQKRYTYGTDKNDVTEAPEIKSL-----KGHLSENIIVILPETITT 462
QY 192 QHNKVTSPKITTAKPINRPSLPPNSDT-----SKETSILVNKETTIVETKETT 240
Db 463 -----TTTTTKPVVLTCPITISPDITPKPSTTAVTKSTPKISSTEQHSITTTAKITTKRPT 520
QY 241 TTNKQSTQCKEKTSAKTSQLEKTSAXDLAPTSKVLAKPTKABTTTKGPAITPKPEP 300
Db 521 VTEKTSATEKPRTTVVTTTQKRSTTHNTSPDTKTITRSTLSPKTTTSPSTTSTT 580
QY 301 TPTTPKEPASTTKEPTTIKASPTTKPEAPTTTKSAPTTKPEAPTTTKEPAPTTPK 360
Db 581 TPST-----TTPSTTTPST-----TTPS-----TTTTPTTTIVK-----VSTHRPTTSOK 621
QY 361 EPAPTTPKEPAPTTTKSAPTTKPEAPTTPKPEAPTTPKPEAPTTTKEPAPTTPK 420
Db 622 -----TIT-----ASTTKTKTTSK-----TTKTIDPTSTTSKLSLTTQKTTTTTHKFTAAT 670
QY 421 -----TKPEAPTTPK-----EPAPTAKPEAPTTPKPEAPTTTKEPAPTTPK 464
Db 671 TSTEKPTTTTTEKTSVTSTTKTSSTESPKSTSTGKPTTTPKPESTTTTTPKVTITTIQI 730
QY 465 SPPTKEPAPTTTKSAPTTTKEPAPTTKSAPTTPKPEPSTTTTKEPAPTTPKPEAPTTPK 524
Db 731 TTTTPLASSTETSTQPPTTTTQPTTTTTLTVTPKSTTTTTTTEKPISTSSPKPTTTTQK 790
QY 525 -----KPAPTTPKEPAP-----TTPKEPAPTTTKEPAPTTPKPEAPTTTPK 563
Db 791 TTSTAENTTKVAITTOKETTTQSTSTTIFTRKTTTNNPEPTSTKPTITSTTPKPSSTTP 850
QY 564 K-STAPTTPKLTPTTPEKLAAPTTPK-----APTPELAPTTPPEPTTTPPEP 614
Db 851 KTSTVASSTETKTISPKPTTEKSTENPTNSVKTSALESSTORASTTSEPTKITT-QNI 909
QY 615 APPTKAAAPNTPKPEAPTTPKPEAPT--TPKPEAPTTPKETAAPTTPKGPAPTTLKEPAP 672
Db 910 TTTTPKPTLLKTSQEAATTSQKSVTVIITTKKATSSPLTILSTEEPNTPKPLRTTTP 969
QY 673 TTPKKPAPKELAPTTTKEPT--STTSKAPAPTTPKGTAPTTPK-----EPAPTTPK 721
Db 970 TTTSVTATTTTITISESTETTSOKPKSTTPTSTRTTPKVTTVIVSTQNTPTTTTSK 1029
QY 722 -----EPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKG-----PTST-----763
Db 1030 TSTVTIITPNPSPSTQPT--TTTQPTSTIT-----ASTISIGTTRITPTTNPQNS 1078
QY 764 -----TSDK-----767
Db 1079 TSSDTLATVTRPPCPDPDSTSDKNNTACTQELQVNNLELQSPKQEQFTHTRTHALT 1138

```

```

QY 768 -----PAPTTPKETAPTTPKEPA-----PTTP--K 790
Db 1139 GSRNLTGQGVDPDMDAPSSAEASQOATKAPTMTSLAAAHLLQKLFHILSTTPSR 1198
QY 791 KPAPTTPETPPPTTSEVS-----808
Db 1199 EHAPTQRPSPQSPSSQSRGVTIAQMARHNLATSKPFIASHRLSLIOQLASTOKRSIPPK 1258
QY 809 ---TPTTTKP-----TTIHKSPDESTPELSAEBTPKALENSPKPGVPTTKTPAAT 857
Db 1259 TLVTHNTTKPESEYDYDSEYTDEN-EVLDKTOPRAMSST-----TVAAV 1307
QY 858 KPMTTAKTKTTERDLRTTPTTTAAPKMTKSTATTTE-----896
Db 1308 LPAYPST-----TTERE-----POKTSSSPSPTKATSTTTQPIETTGDLEVDSSGSDYV 1359
QY 897 -----KTESKITATTQVT-----911
Db 1360 NDANDISGVVNSLEARNFLLSKLORLTQIERTEAKKATSTSTTDDAPKTSSSSTSP 1419
QY 912 -STTTQDTTPPKITTLKTTT-----LAPKVTTTKTIT--TTEIMNKPEE-----953
Db 1420 ASTTSESTSPVSTARSLLTASKHLGPEALSROKSLTPQSAEYVDEDDDDYMEDEPVGSS 1479
QY 954 -----TAK-----956
Db 1480 DAKEKEHGTVLISEKQAAATAKRIAPPQAQPLQAMNLATSEENKVDLKTSDGLQ 1539
QY 957 -----BKDRAT-----NSKATTPKPKPTKAPKPTSTKKPKTM 990
Db 1540 FRRFESSOIPTGEPNPNASIVVAEALSRRADNVGEHLPGTVPHTSDAKAISVSQTTI 1599
QY 991 PRVRKPKT---TPTPRKMTSTMP-----ELNPTSRIAEAMLO---TTRPNQTPNSKLVEV 1040
Db 1600 -RSRESQPKSSSQTPPSISTETPATQAIQNHRTDEATQSTLEIVTQTTPKSAPPATAVPV 1658
QY 1041 NPKSEDAG-----GAEGETPHMLLEPHVFMPEVTPDMDYLPVRVNOGI-----IINP 1087
Db 1659 AVROEVLVSNRPMLLAARNQTVHL-----TPISSIAARAFSNPVSHANRSINP 1707
QY 1088 MLS 1090
Db 1708 LVS 1710

```

Search completed: October 13, 2004, 12:17:01
Job time : 153.845 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 22.2022 Seconds
(without alignments)
4723.889 Million cell updates/sec

Title: SEQ1-B
Perfect score: 5826
Sequence: 1 MAWKLPYILLLLSVFVIQ.....DMDVLPVFNQGIINPMLS 1090

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79: *
1: Piri: *
2: Pir2: *
3: Pir3: *
4: Pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1330.8	22.8	3020	2 A43932	mucin 2 precursor,
2	1127.2	19.3	1664	2 T18262	S-layer protein -
3	1004.7	17.2	1489	2 T31108	cyst germination s
4	1002.8	17.2	1274	2 T16251	hypothetical prote
5	992.5	17.0	7962	2 I38346	elastin titin - hu
6	987.1	16.9	3370	2 T45025	mucin MUC5B, trach
7	974.5	16.7	2387	2 T38226	nascen polypeptid
8	928.3	15.9	1367	1 S48478	glucan 1,4-alpha-g
9	901.5	15.5	3507	2 T34513	hypothetical prote
10	888.9	15.3	1188	2 S49915	extensin-like prot
11	869.3	14.9	1229	2 T25697	hypothetical prote
12	862.5	14.8	6642	2 T29757	protein UNC-89 - C
13	839.5	14.4	2897	2 B48666	cell proliferation
14	839.5	14.4	3256	2 A48666	cell proliferation
15	833.4	14.3	5762	2 A41819	proline-rich pepti
16	822.8	14.1	1151	2 T18535	high molecular mas
17	818.7	14.1	4135	2 T42629	tenascin-X - bovin
18	814.1	14.0	1344	1 A35175	mucin 1 precursor,
19	788.2	13.5	2232	2 T34434	hypothetical prote
20	767	13.2	5862	2 T03454	ALR protein - huma
21	764.4	13.1	1832	2 T31113	mucin-like glycopr
22	761.8	13.1	2142	2 B35098	MHC class III hist
23	752.5	12.9	4006	2 T03070	probable tenascin
24	746.2	12.8	3942	2 T42730	Basoon protein -
25	745	12.8	2225	2 T56063	hypothetical prote
26	744.5	12.8	990	2 I51618	nucleolar phosphop
27	738	12.7	761	2 C84672	hypothetical prote
28	737.7	12.7	4548	1 S06657	apoptosis(a) (EC
29	735.1	12.6	1870	2 S37671	MHC class III hist

MHC class III hist
gene IF3 protein -
hypothetical prote
hypothetical prote
microtubule-associ
Basoon protein -
hypothetical prote
hypothetical prote
ascites siialoglyco
versican precursor
cell proliferation
ALR protein - huma
hypothetical prote
hypothetical prote
tenascin-X precurs
hypothetical prote

30 734.7 12.6 1872 2 S36152
31 733.5 12.6 924 2 S27923
32 731.3 12.6 971 2 T19431
33 730.4 12.5 1777 2 T34369
34 730.2 12.5 2774 2 A43359
35 726.1 12.5 3938 2 T42761
36 716.2 12.3 801 2 T29018
37 715.2 12.3 1611 2 T38236
38 713.7 12.3 1630 2 A53577
39 702.5 12.1 3381 2 T42389
40 696.2 11.9 2938 2 T30249
41 696.1 11.9 4957 2 T03455
42 696 11.9 839 2 F75518
43 694.1 11.9 5105 2 T32650
44 693.5 11.9 3566 1 A40701
45 691.1 11.9 5170 2 T15348

ALIGNMENTS

RESULT 1
A43932
mucin 2 precursor, intestinal - human (fragments)
N:Alternate names: mucin SMUC-41
C:Species: Homo sapiens (man)
C:Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 09-Jul-2004
C:Accession: A49963; A45106; B45106; A43932; B35532; A61257; PQ0329
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A:Reference number: A49963; MUID:94132002; PMID:8300571
A:Accession: A49963
A:Molecule type: mRNA
A:Residues: 1-639 <GU1>
A:Cross-references: UNIPROT:Q02817; UNIPROT:Q14884; GB:L21998
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr
A:Reference number: A45106; MUID:93016075; PMID:1400449
A:Accession: A45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 626-1895 <GU2>
A:Cross-references: GB:M94131; NID:G186395; PIDN:AAA59163.1; PID:G186396
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)
A:Accession: B45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 2037-3020 <GU3>
A:Cross-references: GB:M94132; NID:G186397; PIDN:AAA59164.1; PID:G186398
A:Experimental source: colon
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M.
J. Clin. Invest. 88, 1005-1013, 1991
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp
A:Reference number: A43932; MUID:91358717; PMID:1885763
A:Accession: A43932
A:Molecule type: DNA
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A:Cross-references: GB:M74027; NID:G188863; PIDN:AAA59875.1; PID:G188864
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIP:55750)
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden
A:Reference number: A33532; MUID:89197956; PMID:2703501
A:Accession: B3532
A:Molecule type: mRNA
A:Residues: 1916-2193 <GU4>
A:Cross-references: GB:M22405; NID:G188873; PIDN:AAA36334.1; PID:G188874
A:Experimental source: intestine
R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

J. Clin. Invest. 87, 77-82, 1991
 A:Title: Human bronchus and intestine express the same mucin gene.
 A:Reference number: A61257; PMID:91086481; PMID:1985113
 A:Accession: A61257
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
 A:Experimental source: bronchus
 R:Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Jones, C.; Forstner, Biochem. Biophys. Res. Commun. 183, 821-828, 1992
 A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-terminus
 A:Reference number: PQ0328; PMID:92198477; PMID:1550588
 A:Accession: PQ0328
 A:Molecule type: mRNA
 A:Residues: 2328-2468 <XUG>
 A:Cross-references: GB:M86523
 A:Experimental source: small intestine
 A:Accession: PQ0329
 A:Molecule type: protein
 A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
 C:Genetics:
 A:Gene: GDB:MUC2
 A:Cross-references: GDB:120203; OMIM:158370
 A:Map position: lip15.5-lip15.5
 C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von Willebrand factor type C repeat homology
 C:Keywords: glycoprotein; intestine; tandem repeat
 F:2766-2834/Domain: von Willebrand factor type C repeat homology <VMC>
 Query Match 22.8%; Score 1330.8; DB 2; Length 3020;
 Best Local Similarity 20.5%; Pred. No. 6.4e-19; Mismatches 495; Indels 992; Gaps 63;
 Matches 411; Conservative 108; Gaps 63;
 QY 6 LPYLLLSLVFVIOQVSSQDLSSCAGRCG-----EG-----YSR 40
 DB 502 LQVQLAPVQLFVTLDAQSQ--GVQGLCGNFGLEGDDFTASGLVEATGAGFANTWKA 559
 QY 41 DATCN-----CDY---48
 DB 560 QSTCHDKLDLDDPCLSNIESANYAEHWCSSLKKTETTFPGCHSAVDPAEYKRCYDTC 619
 QY 49 NQHYMEC-----CPDFKRVCTA-----ELSK 71
 DB 620 NQNNEDCLCAALSSYARACTAGVWLWGREHVCNKGVDGSCPNQVFLYNLTTCQCTCR 679
 QY 72 -----GSCFESF-----ERGR-----ECDC-----DAOCKYDK 95
 DB 680 SLSEADSHCLGFPVDDGCGCPDHTFLDEKRCVFLAKCSYHRLYLEAGDVVVRQER 739
 QY 96 C-----96
 DB 740 CVCRDGLHCRQIRLIGOSCTAPKIHMDCSNLTALATSKPRALSCQTLAAGYHTECVSG 799
 QY 97 ---CPD-----99
 DB 800 CVCPLDMDGRGCGWKEKCFVENNDLYSSGAKIKVDCNCTCKEGRWVCTQAVCHGT 859
 QY 100 -----YESFCABEHS-----VSENOESSSSSSSSSS 126
 DB 860 CSIVSGHYITFDGKYDFDGHCSYVAVQDYCGNSSLGSSFIITENVPCGTGVTCSKA 919
 QY 127 STIWKIKSSKNSAANRELQKLKVDNKK-----155
 DB 920 IKIFMGR-----ELKLEDKRVVIOQDEGHVAYTTREVQYLVWESSGTII 967
 QY 156 ---NRT---KKPTKPPVVDAGS---GLDNGDFV-----TT 185
 DB 968 VINDKRTVFIKLAPSYKGTGCLGNFDRHSNNDFTRDMVSVSELDFGNSWKEAPTC 1027
 QY 186 PDTSTTQ-----HNKVTSP-----200
 DB 1028 PDVSTNPBCSLNPHRRSWAEKQCSILKSSVFSICHSKVDPKPPYEAHVDCSDTGGD 1087
 QY 201 -----KITTAKP 207

DB 1088 CECFCSAVASYAQECTKEGACVFWRTDLCPIFCDYNNPHECEWHYEPGCRNSRETCT 1147
 QY 208 IN-----PPPSLP-----215
 DB 1148 INGIHSNISVSVLEGCPYPCPKDRPIVEEDLKKCVTADKCGCVEDTHYPGASVPTBET 1207
 QY 216 -----PNSDTSKE-----TSLVANKET-----232
 DB 1208 CKSCVCTNSQVVRPEBKILNQDGAFCVWEICGPNGTVEKHFNCSITRSTLT 1267
 QY 233 -----TVETRETTTNKQSTDGKETSAR-----258
 DB 1268 FTTITLPTTPTSTTTTTTTTTTSSVLSLTTPKLCCLMSDWINEDHPSSGSDGDPED 1327
 QY 259 -----ETQSIKTSKD-----270
 DB 1328 GVGAPEDIERSVKDPHLSLEHQHGVQCDVSVGFICKNEQDFNGFGLCYDYKIRVN 1387
 QY 271 -----LAPTSKVLAKPTKPAETTTKGPALTTPKEPTTPKEPASTTPKEPTTIK 322
 DB 1388 CCWPMDCITTPSPPTTTPSPPTTTLPTTTPSPPTTTPPTTTPPTTTPPTTTPPTT 1447
 QY 323 SAP-TTPKEPAPTTKSAATTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 380
 DB 1448 PLPTTTPSPPTTTPPTTTPSP-PITTPSPPTTTPSPPTTTPPTTTPSPPTTTPSP 1505
 QY 381 TP-----KBPAPTTKPKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 432
 DB 1506 TITTPASATTLPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTT 1565
 QY 433 APTAPKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 492
 DB 1566 PTTTTTPPTTTPSPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPSP 1624
 QY 493 KSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 551
 DB 1625 ---PTTP---PTSTTLPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTP 1677
 QY 552 PTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 611
 DB 1678 TITTPSPPTTTP-PPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTP 1728
 QY 612 EBPAPTTTPKAAAPNTKBP-----PTTPKEPAPTTTPKEPAPTTTP-----KETAP 659
 DB 1729 TTPSPPTT-----TPSPPTTMTLPTTTPSSPLTTLTPPSITPTTTPSPPTTTP 1781
 QY 660 -----KG-----661
 DB 1782 CVPLCNWTCGLSDGKFNHFKPGDTELICDVGCPGMAANISCRATWYDPVPIGQLGQIV 1841
 QY 662 -----TAPTTLKEPAPTTTPKPKAPKEL 683
 DB 1842 CDVSVGLICKNEDQKPGVIPMAFCLNIEINVQCECVTQPTM-----TTTTENPTPT 1898
 QY 684 APTT-TKEPTSTTSDKP-----APT-TPKGTAPTTTPKEPAPTTTPKEPAPTTTP 736
 DB 1899 TTTTTPPTTTPSTQSNGLQAPPTTPTSTTTTTPPTTTPPTTTPPTTTPPTTTPPT 1957
 QY 737 KEAPTTTPKPKAPKELAPTTTKGPTST-TSDKAPTTTPKEPAPTTTPKEPAPTTTP 792
 DB 1958 PPTPTGTGTPTTTLITTTTTTMTPTPTSTKSTTPTTPTTPTTPTTPTTPTTPT 2017
 QY 793 -----APTTPETPTPTTSEVSTPTTK-----EPTTIHKSDESESTPELSAET 843
 DB 2018 PSTTTTTTPPTTTPGTPPTTSTADIAELITSNPPPESSIPQTSRSTSSPLTES 2077
 QY 844 KBPGVPTTKTAAKPEMTTAKOKTTERDLRTPETTTAAAPKMTKETATTTTEKTESKI 903
 DB 2078 LLSTLPPALEMTSTAPPSTAPTITSGGHTLSPPSTTTPPGTPTRGTTT--GSSSAP 2135
 QY 904 TATTQVSTTQDTPPKITT---LKTTLAPKVTTKTITTEIMNKPEE-----953

Db 2136 TPSTVQTTTSAWTPPTPLSTPSIIRTTGLRPYPSSVLICCVLNDTYVAPGEEVNGTY 2195
Qy 954 -----TAKPKDRATNSKATTPKPKQPTKAPKPKSTKPKTMP 991
Db 2196 GTCYFVNGSLCTLEFYVNSCFSTPPTPSK-STPTSPKSPSTPKPTGTPKPPCP 2254
Qy 992 RVRKPKTTTTPRKMTSTMPELNPTSGRIAEAMLQTTTRNQTNSKLVEVNP----- 1042
Db 2255 DFDPPP-----QENETWLCDFM-AICKYNNVTVEIVKVECEPPMPTCSNG 2300
Qy 1043 -----KSEDAGGA-----EGETHMLLRHRVFMPEVTP 1070
Db 2301 LOPVRVEDPGCWHWECDYCTGWDGPHYVTFDGLYYSQGNCTYVLVE-----EISP 2354
Qy 1071 DMD-----1073
Db 2355 SVDNFGVYIDNYHCDENDKVCERTLIVRHETQEVLLIKVTHMPQVQVQVNRQVVALPY 2414
Qy 1074 -----YIPRVNQGIIN 1086
Db 2415 KKYGLEVYQSGINYVVDIPELGLVLS 2440

RESULT 2
T18262
S-layer protein - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18262
R:Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
e.
A:Reference number: Z18847; MUID:93209931; PMID:8458832
A:Accession: T18262
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1664 <FUI>
A:Cross-references: UNIPROT:Q06852; EMBL:X67506; NID:G296879; PID:G296881; PIDN:CAA47841

Query Match 19.3%; Score 1127.2; DB 2; Length 1664;
Best Local Similarity 21.4%; Pred. No. 2.9e-15;
Matches 369; Conservative 127; Mismatches 427; Indels 799; Gaps 68;

Qy 4 KTLPIYLLLLSVFIQVSSODLSSCAGRCGEGYSDATNCDCVNCQHYMECCP----- 58
Db 6 KVLISILLTLL-----IISTSVNMS-----FAEATPSIEMV 37
Qy 59 -----DFKRVCTABLSC-----GRCFESFERGECDC 86
Db 38 LDKTEVHVGDVITAIKVNIRKLAGYQLNIKFDPEVLQVPDPATGEEFT----- 87
Qy 87 DAQCKYDKCCP-----DYSFCAEEH-- 108
Db 88 -----DKMPVNRVLNTNSKYGTPVAGNDIKSGIINFATGVNNLTAYKSSGIDEHTG 140
Qy 109 -----SVSENOBSSSSSSSSSSSTIW-----KIKSS 135
Db 141 IIGETGFKVLKQNTSIFEDTLSPGALSIGLSLFDWDAEITGYEVIQDPLIVVEAPL 200
Qy 136 KNSAANRELQK- KLVKD-----NKN-----RTKKKT----- 163
Db 201 KDAVALELDKTKVKVGDIIITATIKIENMKNFAGYQLNIKVDPTMLBAILETGSALAKR 260
Qy 164 -----PKPPVDEAGSLDN----- 178
Db 261 TWPVTGGVQSDNYGKTTAVANDVAGIINFAPAEAYSLTKYRETVGVAEETGIIGKIGFR 320
Qy 179 -----GDFKVTTPDT-----STQHNKV-----STSPK 201
Db 321 VLKAGSTAIRFEDTTAMPALGTYWFDYNGENIKGYVWQGEIVAEGBGPEEPTSP 380
Qy 202 ITTAKPINRPSLP-----PNS-----DTSK----- 222

Db 381 VPIETPVDPPTVTEEPVSELPSDYVIMELDKTKVKVGDIIITATIKIENMKNFAGYQLN 440
Qy 223 -----ETSUTVNKETTVETKETTTTNTKQTSDDGKEKTT----- 256
Db 441 IKYDPTMLEALETGSALAKRTWPTGGTV-----LQSDNYGKTTAVANDVAGIINFA 495
Qy 257 -----AKETQSIEKTSADLAPTCKVLAKPTPKAETTT----- 289
Db 496 EAYSNLTXYRETVGVAEETGIIGKIPKVLKAGSTAI-----RFEDTAMPALGEGTYMFD 550
Qy 290 -----KGPALTTP-----KEPTPT-TPKEPASTTPKEPT----- 319
Db 551 WYGENIKGYVWQGEIVAEGBRPTBEPVPTETPVDPTPTVTEEPVSELPSDYVIMELD 610
Qy 320 ----- 319
Db 611 KTKVKGEDVIATIRVNNIKNLGYQIGIKYDKVLEAFNIEGTGDPIDEGTWPAVGTL 670
Qy 320 -----TIKSAPTPK 329
Db 671 KNRDYLPTGVAINNVSKGILNFAAYVYFDDYREEGKSEDTGIIGNIFRVLKAEPTIR 730
Qy 330 -----EPAPTTKSAPTTPKAPAPTTPKAPAPTTPKAPAPTTP 359
Db 731 FEELESMPGSIIDGYMLDYNLRISGYVVIQAPAKAAS-----DEPIPTDPSDEPTPS 785
Qy 360 KEPAP-----TTTKEPAPTTPKAPAPTTPKAPAPTTPKAPAPT-----TP 406
Db 786 DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 845
Qy 407 K-EPTPT-TPKAPAPT-TPKAPAPTTPKAPAPTAPKKA-----PTTPKEPAPT-TPKEPAPT 460
Db 846 SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 905
Qy 461 TKEPSPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAP 520
Db 906 TPSEPTPSDEPTPS-----DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 962
Qy 521 TTPKKA-----PTTPKEPAPT-TPKEPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKE 565
Db 963 SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1019
Qy 566 TAPTPPKLTPTPKAPAPTTPKAPAPT-----TPEELAPTTPKEPTPT-TPKEPAPTTPKA 621
Db 1020 -----EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1061
Qy 622 AAPNTPKAPAPTTPKKA-----PTTPKEPAPT-TPKETAPTTPKAPTTPKAPAPTTPKAP 678
Db 1062 DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1119
Qy 679 APKELAPTTTKEPT-STTSKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPK 737
Db 1120 TPSP-EPSPDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1172
Qy 738 EPAPT-TPKAPAPKELAPTTTKEPTSTTSKAPAPT-----TPKETAPTTPKAPAPT-TPKAP 792
Db 1173 EPTSETPEPT-----PTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1226
Qy 793 APPTPEPTPTTSEVSTPTTKEPTTIHKSPDESTPELSAEPKALENSPKPEPGVPTTK 852
Db 1227 IPTDTPSDEPTPSD-----EPTPSDEPT-----PSDEPTP-SDEPTP-----SEPTPEP-IPDPT 1274
Qy 853 TPAATKPEMTTAKDKTTEADLTPTTETTTAAAPKMTKATTTTEKTESKITATTITQVTS 912
Db 1275 PSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1334
Qy 913 TTTQDTPPTPKITITLKTTLAPKVTITTKGTTTTTINMKNPEETAKPKDRATNSKATPKPKQ 972
Db 1335 TPSEPTPSDEPT-----PSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1369
Qy 973 -KPTKAPKPTSTYKPKTMRVKKPTTPTPKMTS-----T 1008

```

Db      1370  PTPSETPEPT-----PTTPTFTPTTPTSGSGSGSGGGGGGGGTVP 1416
Qy      1009  MPELNPTSRIAEAMLQTTTFNPOTNPSKLVENVPKSBDAGACETPHMLLRPH--VFM 1065
Db      1417  SPTPTPTS-----KPTSTPAETEIE-EPTPSDVFGAIGGEHRAYLRCYPDGSPR 1464
Qy      1066  PE-----1067
Db      1465  PERNITRAEAAVIPAKLIGADESYGAQASPSYSLADTHWAAWAIFATSQGLFKGYPDG 1524
Qy      1068  -VTPD-----MDVLPYRPNQGI-----IINP 1087
Db      1525  TFKPDQNIITRAEFATVVLHFLTKVKGEIMSKLATIDISNP 1565

RESULT 3
T31108
cyst germination specific acidic repeat protein precursor - Phytophthora infestans
C;Species: Phytophthora infestans (potato late blight agent)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C;Accession: T31108
R;Goerthardt, B.
submitted to the EMBL Data Library, April 1998
A;Reference number: Z20986
A;Accession: T31108
A;Status: Preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1489 <G0E>
A;Cross-references: UNIPROT:O96449; EMBL:AF061185; NID:G3851513; PID:G3851514; PIDN:AACT
C;Genetics:
A;Gene: car90

```

[illegible]

RESULT 4

T16251

110231 hypothetical protein F35A5.1 - *Caenorhabditis elegans*

C: species: *Caenorhabditis elegans*

C/Species: Caenorhabditis elegans
C/Date: 20-Sep-1999 #sequence revision 20-Sep-1999 #text_change 09-Jul-2004

C:Accession: T16251
R:Reimbach, D.
A:Description: The EMBL Data Library, January 1996
submitted to the sequence of C. elegans cosmid F35A5.
A:Reference number: Z16485
A:Accession: T16251
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1274 <LEI>
A:Cross-references: UNIPROT:Q20007; EMBL:U46675; NID:g1166621; PIDN:AAB526
A:Experimental source: strain Bristol N2; clone F35A5
C:Genetics:
A:Gene: CESP:F35A5.1
A:Map position: X
A:Introns: 1272/2

Query Match
Best Local Similarity 17.2%; Score 1002.8; DB 2; Length 1274;
Matches 330; Conservative 99; Mismatches 309; Indels 692; Gaps 75;

QY 158 TKKKTTP-----KPP-VVDEAGSLDNGDFKVTTPDSTTOHNVKSTSPKITTAKP 207
DB 2 SRAPPTPIKNPAKKWPPWESVDE----- 25

QY 208 INPRSLPNSDTSKETSITVNKETTIVETKETTITNNKOTSDGKEKTTSAKETQSIKTS 267
DB 26 -----EEMVEVDEETPAFSK-----LEKPSLR----- 49

QY 268 AKDLAPTQKVLAKTPKA-----ETTKGPAITTP-----KEPTP- 302
DB 50 -KD-AFTKPV-----PSFGAPSPVPIKNPVKKWAPWEDDEPBEAAPVPAKKVRDPSK 104

QY 303 TTPKSPASTPK-----PTPT-----TIKSAPTTP 328
DB 105 KVPKPRDASPKKIWAKEPBTLPVPPVKNVKKFKAPWEDDEVDVDDKADAPVP 164

QY 329 -----KEPA-----PTTKSAPTTPKEP-----APTTPKEPA-----PTT 358
DB 165 AKKTPVLKKEPAAAKPRDPSKKAAPSKKHDPVPPPTPIKNPAKKWKPWEDDEVTE 224

QY 359 P-KEPAPTTPKEPA-----PTTKSAPT----- 380
DB 225 EIKPEPATRKVPALKKKEPSTSVKPVSDPSPTTKKVPVKEPVEPPTPIKNPTKKWKPW 284

QY 381 ---TP-----KEPAP-----TTPKPAAPTTPKEPAPTTPK 407
DB 285 EDETVEEVKEPVPVEKAPVLLKKDPAPAKARDSPSKAAPKVEPSSPVP-PTPVK 343

QY 408 EPT-----PTTKGPAPTTPK-----EPAPTTP-KEPAP----- 434
DB 344 NPVKYKPPWEVDDEPABEVKPSAPEKKTPLKKEPEPSSPTPSDPSPKAAPAVK 403

QY 435 ---TAPKAP-----PTTKGPAP-----TTPKEPAPTTP-- 453
DB 404 ROSSPKKATPLOADPKAOEVPTPVKNPVKKYKPPWEVDDEPVEVQEPAPAKKTPVL 463

QY 454 ---KEPA-----PTTKKEP-----SPTTP-KEPA----- 473
DB 464 KRKEPAAKDTAKPATSKTPTETPEKDPVKPRDSSPKKVAAPDSQAAPATPVGNPVKWR 523

QY 474 -----PTTKSAPT--TTKEPAP-----TTTKSAPTTPKEPSTTTKE 509
DB 524 PPWEDETPADDVSKPTDAKKTPLAKKDPAPAKESLRPKADTK-APAKPRDPSP---KK 579

QY 510 PAPTTPKEPAPTTPK-----PAPTTPKEPAPTTPKEP-----APTTPKK 549
DB 580 VAPTAPKKTPLVLAKEPAGPADSKTKEPKSKPRDPSFKKAVPAKVPKTEVAPAAVK 639

QY 550 PAP-----TAPKEPAPTTPKETATTP-----KKLTP-----TTPKEKAP 584
DB 640 PEPISKPDATAPKAEPSNP--WSPPTPVKNPVKKWKPWEDDDAPAKPVSLPEPEKKT 697

QY 585 TTPKEPAPTTP-----ELAPTTPKEPPT--TTPKEPAPTTP----- 618

DB 698 VLAKK-APTKPDSEAAADPVSGPSSKDPKLAKKAPVKPRDPSPMKAVP-IKPAKTEVPPA 756
QY 619 -----PK--AAAPNTPEKAPTTPKEP----- 638

DB 757 VVKPBPVAKSRDPSPKKAKAEFNSPVP-PTPVKNPVKKWKPWEDDDAPAPVAVPEP 815
QY 639 -----APTTPKEPAPTTPKEPAPTTPKGTATTTLKBPAPTTP-PKKAPKE----- 682

DB 816 EKKTTPVLAKKTPVKPRDPSPKKAVPAKPSKTTDAPPVSVKVPKPEPSPKAEFNS 875
QY 683 --LAPTTPKEPT-----STSDKPAAPTTPKGTAPT--TPKEP-----APTTPKEP 723

DB 876 PVVPTPVKNPVKKWKPWEDDDDEPTEEVKPSPEKKTPLVLAKEPEKPKDAPKVAAP 935
QY 724 APTTPKGTAPTTLKEPAPT-----TPKKPAP-----KELAPTTPTKGPTS-----TTS 765

DB 936 RDPSPKAVPE--KEPAKAAKPRDLSPKKAIPANTQEAAPTVPKQPVKKWKPWEDD 993
QY 766 DKP-----APTTPKET-----APTTPKEPAP-----TTPKPAAPTTPPETPPTTSEVSTP 810

DB 994 DEPAEPVSAPEPEKKTPLVLAKEPAPRDPSPKKAAPVAAKPDPKTPEVP-----P 1044
QY 811 TTTKEPTTIHKSP-----DESTPELSA-EP---TPKALENSPKPEGV-----P 849

DB 1045 TPVKNPVKKWKPWEDDDDEPSEPVSAPEPEKKTPLVLAKEPAPTTPKPDSEAAADPVSGP 1104
QY 850 TTTTPAAT-----KPEMTTAKDKTTERDLRTTPETT---TAAPKMTKETATTTTEKTES 901

DB 1105 TSDPKLSKKAPVEKPKPTTDPKDDKLKPSPAKKPEKAPPEAPAKKWKPVWDDDPDEPA 1164
QY 902 KITATTTVSTTQTOTTPPKITTLKTTLLAPVTTTKTITTTTEIMNKDEETAKPKDRA 961

DB 1165 DFTVPAPSKPDPDIEDPADPLG-----GPKTKDPK-----LNKKAPAEKTEK- 1206
QY 962 TNSKATTPPKQKPTKAPKPTSTKPKTTPRVRKPK-----TTPTRKMTSTM 1009

DB 1207 -----PKEKEVSKPEPKTEPPKP-AAPKWKPPWEDDDPEPEADFTMPAPKK----- 1253
QY 1010 PELNPISRIAEAMLQITTRNQTPNSKLVENVKPSD-----AGARGETP 1055
DB 1254 -----PTDEDPADSLGGPKPKDP 1271

RESULT 5
138346
elastic titin - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004
C:Accession: I38346
R:Labeit, S.; Kolmerer, B.
Science 270, 293-296, 1995
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330; PMID:7569978
A:Accession: I38346
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-7962 <RES>
A:Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g101
C:Genetics:
A:Gene: CDB:TIN
A:Cross-references: GDB:127867; OMIM:188840
A:Map position: 2q31-2q31

Query Match
Best Local Similarity 17.0%; Score 992.5; DB 2; Length 7962;
Matches 405; Conservative 108; Mismatches 429; Indels 2145; Gaps 95;

QY 4 KTIPIYLLLLSVFVQVSSQDLSSCAGRCGSGYRDATCNC----- 46
DB 4815 KNPFDFIRLAPDAV-----VGESADFECHVTGTQPIKVSWAKDSREI 4858
QY 47 -----DYNQHYNECCPDFKRVCTAELSCKGRCP-ES 77

4859 RSGGYQIYSELENSAHLTVLVKVDGSGQYTCYAVNEVGKD---SCTAQLNIKERLIPPS 4915
Db
78 BEERGECDDAQCKYDKCCPDYSSFCABEHSVSENQESSSSS---SSSSSTI 129
QY
4916 FTK-----RLSETVEETEGNSFKLEGRVAGSQITVA 4947
Db
130 WKIKSSKNSAANRELQKLVKNDKKNRTK----- 159
QY
4948 W---YKN---NIEIQPTSNCEIIFPKNTLVLOVRKAGMNDAGLYTCVSNDSALCTS 5000
Db
160 ---KKPTPKPVUDE-----AG----- 173
QY
5001 SIVIKER-KKPPVFDQHLTPVTVSEGEYVQLSCHVQSGSEPIRIQWLKAGREIKPSDRCSF 5059
Db
174 ---SG-----LDNGDP-----KVTPDSTTQHNNKVSTSPKITTAKPINRPSL 214
QY
5060 SPASCTAVLELRDVAKADSGDYVCASNVAGSDT-----TKSKVT---IKDPVAV 5106
Db
215 PPNSD-----TSKETSJLVNKT----- 233
QY
5107 APATKKAADVGRLEFPVSEFQIRVVVEKTTATFIKAVGGDPPIPNVKWTKGWRLQNOGRV 5166
Db
234 ---VETKETTIN-----KOTSDG----- 250
QY
5167 FIHQGDEAKLEIRDTTKTDSGLYRCVAFNEHGEIBSNVLQVDERKKQKIEGDLRAML 5226
Db
251 -----KEKTT 255
QY
5227 KKTPIKXGAGBEEIDIMELLKNVDPKYEKYARMYGITDFRGLLOAFELLKQSEET 5286
Db
256 SAKETQSTEK-----TSAD----- 270
QY
5287 HRLEIEIERSEDRDKFEELVSFIOQLSOTEPVTLIKDIENQTVLKNDNAVEFIDIKI 5346
Db
271 -----LAPTSK----- 276
QY
5347 NYPEIKLSWYKTEKLEPSDKFEISIDGDRHTRLVKNCQLKQGNVLCVGPHTASAKLT 5406
Db
277 ----- 276
QY
5407 VIEPAWHERLQDVTLEKQGTCTMTVQFVSPNVKSEWFRNGRILKPOGRHKTEVEHKVHL 5466
Db
277 ----- 276
QY
5467 TIADVRAEDQGYCKYEDLETSALRLAEPIQFTKRIQNIUVSEHQSATFCEVSFDD 5526
Db
277 -----VLAKPTPKAETTTKGPA 293
QY
5527 AIVTWYKGTBELTSQKYNFRNGRCHVMTTHNVTPDDEGYSVIARLEPRGEARSTAE 5586
Db
294 LTTPE-----TPT----- 307
QY
5587 YLTKEIKLELKPDPIDPSRVPIITMPIRAVPPEIIPVAVPVPVLLLPTEPKPPPKR 5646
Db
308 -----PASTTPKE-----PTPTTI----- 321
QY
5647 IEVTKAVKDAKKVAKPKEMTPREEIVKPPPPPTTLIPAKAPEIIDVSSKAEVVKMT 5706
Db
322 ----- 321
QY
5707 ITRKKEVQKEAEVYKQAVHKRVFIESFEEPYDELEVEPYTEPFQPYVEEDDY 5766
Db
322 ----- 321
QY
5767 EEIKVEAKKEVHEEWEEDFESGQYBYREGEYDEGESEWEAEYQEREVIQVQEVVEESH 5826
Db
322 -----KSATTPK----- 329
QY
5827 ERKVPKVPKAPPPPKVVKVIEKTSRREMEERKQVTVKVPVSKKIVPQKPSRT 5886
Db
330 ----- 329
QY

5887 PVQEEVIEVKVPAVHTKKMVISSEKMFASHTTEESVTVPEVQKEIVTEEKIHVAVSKR 5946
Db
330 -EPAPTTTKSAPTTPKEPA-----PTTTKEPAPTTPKEPAPTTPTTKEPAPTTPK 379
QY
5947 VEPPP---KVPELEKAPAEVAVPVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 5997
Db
380 TTPKEPA-----PTTPKPKAP-----TTPE 400
QY
5998 VPKEFAAPPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 6057
Db
401 --PAPTTPKEPTTPKE--PAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 442
QY
6058 ESPPPAVPSIPKKKVPKEERKVPKKEEVPKPPKVPKVPKVPKVPKVPKVPKVPKVPK 6115
Db
443 ----- 442
QY
6116 PPPRAVSKTIVVEKRFVAEEKLSFAVPQVTRHEVSAEEEWSYSEEEBEGVSI SVR 6175
Db
443 -----TTPKEPA 449
QY
6176 BEEREEEAETVEYVMEPEEYVVEEKLHIISKRAEAPAEVTEROEKIVLKPPIA 6235
Db
450 -----PTTPKE-----PAPTTPKE--PSPTTPKEP--APTTPKAPTTP--KE 486
QY
6236 KIEEPPPAKVPKAPKIVPEKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPK 6295
Db
487 PAP-----TTKSAP-----TTPKBSPTTKB-----PA 511
QY
6296 PLPAKVTEKHMOTQEBKVLNVAVTKEAPPKARVPEEPKRAVPEEKVLKLPKREEP 6355
Db
512 PTT-----PK-EPAP----- 520
QY
6356 KVTEFRKRVVKEEKVSIAPKREPOPIKEVTIMEKERAYTLEBEAVSQOBEYEE 6415
Db
521 -----TTPKAPPT--TTPKEPAPTTPKE-- 541
QY
6416 YDYKEFEYEPTEDYQVEEYEEYEEYEEYEEYEEYEEYEEYEEYEEYEEYEEYEEYEE 6475
Db
542 -----PAPTTPKAPTAPEP----- 558
QY
6476 PAKVLKXAVPEEKVPVPIPKLKPVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 6535
Db
559 -----APTTPKETAP----- 568
QY
6536 PMKPRVVAEEKVPVPRKEVAPPVPRVPEVKLEPEEVAPEEVEVTHVEEYLVVEE 6595
Db
569 -----TTPKLTPTTPPEKLAPTTPKEP-----APTPELA 599
QY
6596 HEEEFITIEEVEVPIPVKV-PEVPRKEVPBEKVPVVPKKEAPPKVPKVPKVPKVPK 6654
Db
600 PT-----TPEEPTPTTPKEPAPT-----TPKAAAPNTPKPEPAPTTPKEP 640
QY
6655 FVLIPEKKEPPPAKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 6713
Db
641 T-----TKPEPAPTTPKETAPTTPKGT--APTTPKEPAPTTPK----- 677
QY
6714 APKVEAPPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 6773
Db
678 ----- 677
QY
6774 FEEIIVPEEVLPEEVEVLPEEVEVLPEEVEVLPEEVEVLPEEVEVLPEEVEVLPEE 6833
Db
678 -----PAP-----KELAP----- 685
QY
6834 PEEEVLPEVKVPVPAVPEIKKVTEKKVVPKKEAPPAKVPKVPKVPKVPKVPKVPK 6893
Db
686 -----TTPKEP-----TSSTSKPAPTTPK 706
QY
6894 EEEVLPEVEEPEEPEEPISEEEIPEEPPSIESEVEEAVPRVPEVIEKAVPEAPTPVK 6953
Db
707 TAPTTP-----PKE--PAPTTPKSPAP----- 725
QY
6954 EAPPKAVSKKIPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPK 7013
Db

QY	726	-----	725
Db	7014	EKVSVAFRQVVVKKERLELVVEAEVBEIPBEEFHEVEEYFEBEGFHEVEEFKLSQHR	7073
QY	726	-----	739
Db	7074	VEEHRVVKHVRVIEVFAEEVEVEKPKAPKGPSEIKIIPKPKPTKVPRKEPPAK	7133
QY	740	APTTPKK-----	781
Db	7134	VVEPKKIVVEKVKVVEPRVPPTKV-----	7183
QY	782	KEPAPTTPKKAPTTP-----	826
Db	7184	EAPPKVPEAPKEVVEKVPVPPPKPEV-PPTKVPEVPKAAVPEKKVPEAIPPKPEP	7242
QY	827	TPELSAEP--TPKALENSPKERGVTT-----	869
Db	7243	PPVEVEEPEEPEPSA---PPKPEVPPVPEVPEKVPKVPAPPAAPPKPEV-----	7291
QY	870	TERDLRTPTTPTTAAPKMTKEATTTEKTTESKITTATTQVTTSTTQDTTFFKITTLLKT	929
Db	7292	-----TPVKVPEAPK-----	7327
QY	930	TLAPKVTTTKTITTT-----	955
Db	7328	---PKVAVPEKKVPEAIPPKPEPPPEVPEEPEEVALBEPDAEVVEPEPAAPQVTVPP	7384
QY	956	-----	961
Db	7385	KNPVEKAPAVAKKPELPVVKVPEVPEVVEKKVVLVPPKPEAPPKVPVPEV	7444
QY	962	TNSKATTPK-----	987
Db	7445	PEKKVAVPKPEVPPAKVPEVKKVLEBKPAVPVPERAESPPPEVPEEPEEIAPEEBIA	7504
QY	988	-----	1000
Db	7505	PEEKEFPVVAEEBEVEVPPPAVPEBPKKIIPKPKVPEVVKKPEAPPKPEPEKVIEXPKL	7564
QY	1001	-----	1007
Db	7565	KRPPEPPPPAPPKEDVKEKIFQKAIKPKKVPENQVPEKVELTLKVPKGKKVKRLLP	7624
QY	1008	-----	1028
Db	7625	ERKPEKBEWLKSVLRKPEBEEKPEKLEKVKKPAVPEPPPPKPEVEVEVPTVKR	7684
QY	1029	-----	1070
Db	7685	ERKPEPKVPEIKPAIPAPPEPKPE-----	7736
QY	1071	-----	1083
Db	7737	VPVVGKKAEPKAEAAKPKGPIKGV	7763
RESULT 6			
T45025			
mucin MUC5B, tracheobronchial [imported] - human (fragment)			
C:Species: Homo sapiens (man)			
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000			
C/Accession: T45025			
R;Desseyn, J.L.; Guyonnet-Duperrat, V.; Porchet, N.; Aubert, J.P.; Laine, A.			
J. Biol. Chem. 272, 3168-3178, 1997			
A;Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat			
A;Reference number: Z22899; MUID:97166151; PMID:9013550			
A;Accession: T45025			
A;Status: preliminary; translated from GB/EMBL/DDBJ			
A;Molecule type: DNA			
A;Residues: 1-3570 <DES>			
A;Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503			

A;Experimental source: placenta			
C;Genetics:			
A;Gene: MUC5B			
Query Match			
Best Local Similarity 16.9%; Score 987.1; DB 2; Length 3570;			
Matches 435; Conservative 108; Mismatches 445; Indels 2149; Gaps 98;			
QY	23	SSODLSACAGCG-----	44
Db	172	SAGTTTCQRCWTEWDEDEPKSGLGDDVSYDKIRAAAGHLCCQPKDIEQAESPP	231
QY	45	-----NCD-----	59
Db	232	NWTLAQGVQKVDHVFGLVCRNWEQGVFQMCYNIIRVLCDDHDCRGRATTPPTTE	291
QY	60	-----	59
Db	292	LETATTTTQALFSTQPTSSPOLTRAPPASTTAVTFLSEGLTSPTYTSLGTATTGGPR	351
QY	60	-----	59
Db	352	QSAGSTEPTVGVATSLPTRSALPGTGLTWRPSPPTLAPTTMATSRARPTGTAST	411
QY	60	-----PKRVCTAELS-----	74
Db	412	ASKEPLTSLAPLTSELSTQAEIETPRTEMTMSPLNTTTSGQTRCQPKCEWTEWFD	471
QY	75	-----PESFER-----	89
Db	472	VDFTSGVAGDMETFEENIRAAAGKCMWAPKSIETRAENYEVSIQVGOVLTCSETGL	531
QY	90	-CKYDK-----	133
Db	532	TKNEDGTGRNCFNMYNVRVLCDDY-SHC-----	583
QY	134	SSKNSAANRELQKLVKDNKNRTKKPKPPVVDVDEAGSLDNGDFKVT--TPDTSIT	191
Db	584	-----TKPTTATTATGCTATASSTQATAGPHVSTT	617
QY	192	QHNKVTSPKITAK-----	226
Db	618	-----ATTPTVTSKATPFSSPGTATAPALRSTATTPTATSTAIPTSSSLGTTWRLSQ	672
QY	227	-----TVNKETTVEKET-----	273
Db	673	TTTPMATSTATSPETVHTSTVLTTTATTGATGSAVATPS-----	722
QY	274	TSKYL-----	281
Db	723	TTKVLTTTGTATPSSSPGRARTLPVWISTTTPTTRGSTVTPSSIPGTHHTPTVLTT	782
QY	282	-----TPKAETTTKG--PALTTPKETPTT-----	319
Db	783	TTTTVATGSMATPSSSTQSTGTPPSLT-----	838
QY	320	TIKSAP-----	343
Db	839	TTATTPAATSVTPSSALGTTHTPPVNTTATTHGRSLSPSSPHTVCTAWTSATSGILG	898
QY	344	-----KEPAPTTPKEPAPT-----	381
Db	899	TTHTPESTGTSHTPPAATTTGHTSTPALSSPHSSRTTESPPSGTTPPCHTTATSRIT	958
QY	382	-----PKEP--APTT-----	389
Db	959	ATATPKTRISTLLPSQTSAPITVVTMCEPCQANSEWLDYSYFNPBGSGGDFDYSN	1018
QY	390	-----PKKP-----	393
Db	1019	IRAAAGVACEOPLGLECRAQAQPGVPLRELQGVVECSLDGLVCRNREQVQKFKMCFNYE	1078
QY	394	-----APTTPKPAPTTPKE-----	417

Db 1079 IRVFCNCGHCPSTPATSTATSSPTGGTWTWLTTLTATTTTSTGTAFTSLRTP 1138
QY 418 AP---TKKPAPT---TPKEPAPTAKKA---PTTPK---EPAP----- 450
Db 1139 PKVLTTTATPTVSSKATPSSPGTATAPALRSTATTPTATGVTPIPSSSLGTTWR 1198
QY 451 ---TPKEPAPTTPKEPSP-----TPKEPAPTTPKAPATTTKEPAP 489
Db 1199 LSQTTTPTATMTATPSSPTPEAHSTVLTAATTTGATGSAVATPSSIPGTAHTTKVPT 1258
QY 490 TTT---KSAPTTPKEP-----SPTT----- 506
Db 1259 TTTGTATPSSPGTALTTPVWISTTTTPTTRGSTVTPSSIPGTTTATVLTTLTTTAT 1318
QY 507 ---TKKPAPTTPKEPAP---TPKEPAPTTPKEPAP---TPKPAPT--- 529
Db 1319 GSNATPSSSTQSGTPSSLTTTATTTATGATGNTNPSPTGTRPDPVLTATTATPATSS 1378
QY 530 ---TPKEPAPTTPKEPAPT--- 547
Db 1379 TVTPSSALGTTTPPVNTTATTHGRSLSPSPSHVTAWSATGSLGTHITPESRGT 1438
QY 548 ---KKPAPT---APKEPAPTTP---KETAPTTPK--- 573
Db 1439 SHTPAATGTTQHSALSSPHSPSTTESPPSGTTPGHTTATSRATATPSKRTS 1498
QY 574 -LTPTPEKLAPTTP----- 588
Db 1499 TLLPSSPTS-APTITVWGCEPQCAWSEWLDYSYMPGPGSGDFDTYSNIRAGGAVCE 1557
QY 589 KP----- 590
Db 1558 QPLGLECRAQOPVPLRELQGWECSDFLGVCNREOVGFKFCNCFYIRVFCNNGH 1617
QY 591 APTTPEELAPTPEE-----PTTTPKEPAP-----T 617
Db 1618 CESTATSTATPSSPTGTTWLTTEQTAATTATTGTAIPSSPTGTAAPPKVLTSQAT 1677
QY 618 TP---KAAPNTPK-----EPAPTTP-----PKP--- 638
Db 1678 TPTATSSKATSSSPRTATTLPLVLTSTATSKSTATSTFTPIPSSTLGTGTSQNRPPHMA 1737
QY 639 ---APTTKEPAPTTPKEPAPTTP-----PK 660
Db 1738 MSTHPSSTPETHSTSVLTATKATTTATSSMTSPSTPGTWTWLTTLTAATTALPH 1797
QY 661 GTAPTTP-----LKEPAPT-----T 674
Db 1798 GTPSSPTGTTWLTPESTTATVTPGTSTATASSTRATAGTLKVLTSATTPTVISSRAT 1857
QY 675 PKK-----PAKELAPT----- 686
Db 1858 PSSPGTATAPALRSTATTPTATSVTAIPSSSLGTAWRLSQTTTPTATMTATPSSSTP 1917
QY 687 ---TTKEPTSTT----- 695
Db 1918 ETVHTSTVLTTLTATTRTGSVATPSSPTGTAHTTKVPTTTTGTATPSSSPGTALTPPV 1977
QY 696 ---SKPAPTTP----- 703
Db 1978 WISTTTTPTTRGSHVTPSSIPGTTTATVLTTLTTTATGSMATPSSSTQSGTPPSLTT 2037
QY 704 ---PKGTAPTTPKEPAPTTP----- 719
Db 2038 TATTITATGNTNPSPTGTPPIPVLTTLTATTPAATSSVTPSSALGTHITPPVNTTA 2097
QY 720 ---PKP---APT-----TPKGTAPTTP-----LKEP 739
Db 2098 TTHGRSLPSSPHVTPVTAWSATSGILGTHITPSTGTSHTPAATGTTQPSPALSSP 2157
QY 740 AP---TPPKPAPKELAPTTPKGTSTTSKPAPTTPKEP-----AP-----T 779

Db 2158 HPSSRTTESPPSP---GTTTPGHTRGTSRTTATATPSTKTRTSTLLPSSSPTSAPIITVTV 2213
QY 780 TPKEP----- 784
Db 2214 TGCEPOCAWSEWLDYSYMPGPGSGDFDTYSNIRAGGAVCEQPLGLECRAQOPGVPLR 2273
QY 785 ---APTPKAPAPTTPPTTP----- 802
Db 2274 ELQOVVECSLDPLGVCNREOVGFKFCNCFYIRVFCNNGHCPSPATSSATPSSTG 2333
QY 803 TT----- 804
Db 2334 TWTILKLTATTTATTTSTGSTATPSSSTQGPAGTPHVSTTATPTVTSKATPSSPGTA 2393
QY 805 ---SEVSTPTTT-----KEPTTHKS--- 822
Db 2394 TALPALRSTATTPTATSTFTAIPISSSLGTTWRLSQTTTPMATMSTATPSSPTVHTSTV 2453
QY 823 --- 822
Db 2454 LTTTATTTGATGSAVATPSSPTPGTAHTTKVPTTTTGTFTVTPSSSPGTAARTPPVWISTTTT 2513
QY 823 ---PDEST----- 827
Db 2514 PTTSGSTVTPSSIPGTHPTVLTTLTTPQVATGSMATPSSSTQSGTPPSLITTATTITA 2573
QY 828 ---PELSAEPKALEN----- 841
Db 2574 TGSSTPSSPTGTPPIPELTTTATTPAATSSVTPSSALGTHITTPPVNTTATTHGRSL 2633
QY 842 SPKEP-----GVPTTKTPAATKPEMT--- 862
Db 2634 SPSSPHVTAWSATSGTLGTHITPSTGTSHTPAATTTGTTTTSTALSSPHSPSRRT 2693
QY 863 ---TTAKDKTTE----- 871
Db 2694 ESPPSPGTTTPGHTTATSTTATATPSTKTRTSTLLPQSTSAPIITVTVTTCGPOCAWSE 2753
QY 872 --- 871
Db 2754 WLDYSYMPGPGSGDFDTYSNIRAGGAVCEQPLGLECRAQOPGVPLGELQOVVECSLD 2813
QY 872 ---RDLETT 877
Db 2814 FGLVCNREOVGFKFCNCFYIRVFCNNGHCPSPATSSSTAMPSSPTGTTWLTTLTTLT 2873
QY 878 PETT---TAAPKMTKETA-----TTTEKTTESKI-----TATTQV-TSTTT 915
Db 2874 ATTTASTGSTATPSSPTGTAAPPKVLTSPTATPTTATSSKATSSSPRTATTLPLVLTSTAT 2933
QY 915 QDT---TPFKITTLKTTTLAPKVTITK-KITITTEINMKPEE-----TAKPKDRATN 963
Db 2934 KSTATSVTPIPSSILGTTGLPEQTTTPVATMTSTIHPSSPTETHTSTVLTATKATRATS 2993
QY 964 SKAT-----TPKPK----- 973
Db 2994 SSTSSTPPTGTTWLTTLTTLTATTTAGTGTATPSSPTGTTWLTTLTTLTATTTASTGST 3053
QY 974 ---PTKAPKPTS 983
Db 3054 AFLSSTPPTGTTWLTPESTTATVAPPGSTATASSTQATAGTPHVSTTATPTVTSKATP 3113
QY 984 TKPKPT---MPEVRKPKTTPPRKMTSTMPBLNPTS-----RIAEAMLOTTRPNQTP 1033
Db 3114 SSSPGTATAPALRSTATTPTATSTFTAI-----PSSSLGTTWRLSQTTTPTATMTATP 3168
QY 1034 NSKLVEVNPKSEADG-----GAGETPHMLLRHVFMPEVTPDMOYLPRVP----- 1079
Db 3169 SS---TPETVHTSTVLTTLTTLTATTTGATGS-----VATPSSPTGTAHTTKVPTTTT 3214
QY 1080 ---NOGIILNP 1087
Db 3215 TGFTAIPSSSPGTALTP 3231

RESULT 7
T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30826
R:Yotov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A:Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle
A:Reference number: Z30899; MUID:96312450; PMID:8698236
A:Accession: T30826
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2187 <YOT>
A:Cross-references: UNIPROT:P70670; EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AA8187
C:Genetics:
A:Gene: Naca
A:Map position: 10
A:Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A>Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding activ
C:Keywords: alternative splicing; DNA binding; transcription factor
Query Match 16.7%; Score 974.5; DB 2; Length 2187;
Best Local Similarity 21.6%; Pred. No. 4.4e-12;
Matches 364; Conservative 134; Mismatches 419; Indels 765; Gaps 72;
QY 4 KTLPIYLL-----LLSVFIQVSSQDLSSCAGRCBGYSR 40
DB 479 KNLPISALVNVGAPVSPAQAAGLPTKRKDTTLOFLAPIALKESPPSSQASLL-----EVLSE 533
QY 41 D-----ATCNCYDNCQHYMECCPDFKRVCTAELSCCKGCFESFE 79
DB 534 DVTVKTTGGPAPVVRPAIGVAT-----TTSL----- 561
QY 80 RGRECDCAQCKYDKC----- 97
DB 562 ---RADSPAVIRADSCVSPNTVSQLKRSVTDPAAPRTAKNTAPSTTSPLVPLASEGC 618
QY 98 PYEFCABEHSVSNQSSSSSSSSSSSTTWIKSSKNSSANFELOKLVKDNKQR 157
DB 619 P-----VASSMALSPONASVETAL-----ALSPEIKRSV----- 648
QY 158 TKKKPTKPPV-----VDEAGSLDNGDFKVTTPDTSTTQH----- 193
DB 649 ----PFPDPLAEISFNARKVDVSHMESSGSSRQGHDPASVTAKGTWVCLADSLDTS 704
QY 194 -----NKVSTSPKAITAKPINRPSLP 215
DB 705 VSASKGSALSGASSPLYFLEVSFLPEAGLAVQGPKGSLNKLSPTPPSKGAPV-PSTGAP 763
QY 216 PMSDTSKETSLLVNKETTVEKE-----T 239
DB 764 P-----SPKCAPIVPTESSISSQVFAEILPSPQKTPVETASSELISAVOSPKVDPTMSDVT 819
QY 240 TTNKQSTSDGKEKTTSAKETOSI-----EKTSAKDLAPTS---KV 277
DB 820 PTPSPKTSATAVPKDQTSATLTLXSLXPAVTSLSPPKAPVAPSNEATIVPTIPTSLKNA 877
QY 278 LAKPTPKAETTT-----KG-PALAT-----PKE- 299
DB 878 LAAATPKETLATSIPKVTSPSPQKTPKSVSLKGAFAMTSKKATEATAASKDVSPSQFPKEV 937
QY 300 -----PTPTPKPASIT-----PKEPTPTTIKSAPTPKE-PAPITTKSAPITTP 343
DB 938 PLLQHVPTSPSPKSPVSDTLSCALTSPPPKGP-PATLAETPTYPKSPKPAASKKTPATP 996
QY 344 K-----EPAPITTKEPAPITTKEPAPITTKEPAP----- 372
DB 997 SPEGVTAVPLEIPPCSKAPKTAAPKESATSSSKRAPKTAVSKGIVTAVPLEISL 1056

QY 373 ---TTTKSAPTTKBPAPTTPKK-PAPTTPKK-----PAPTTKBPETP-- 411
DB 1057 PLKETSUSA--TPGEKSASSPKRSPKTAGPKETPPGCVGTAVPEISLPLPKETPQONATPNE 1114
QY 412 -----TTPKE-----PAPTTKPE-----A 425
DB 1115 SLAASSQKRSPTSVFKETPPGCVGTAMPLEIPSAPOKAPKTAVPKQIPTPEDAVTILLAGS 1174
QY 426 PTTPKBPAPTAAPKBPAPTT-----PAPTTKPE-----KEPAPT 451
DB 1175 PLSPKKASKTAAAPKEAPATPSVGVIAVSGEISPSPKKSTKTAAPKENSATLPPKRSFKTA 1234
QY 452 TPKEPAPTTTKE-----PSPTTPKE---PAPTTTKSAPTTTKEPAPTTTKSAPTT 498
DB 1235 APKE-TPATSSSEGVTAVPSEISPSPTPASKGVVTLTPKGAPNALAE-SPASPKKVPKT 1292
QY 499 --PKES--PTTKBPAPTTKBPAPTTKBPAPTT-----PKEPAPTTKBPAPTT 546
DB 1293 AAPEETSTTPSPQIKPVKAPKEASATPPSKKTKPTAVPKETSAPSEGVTAVPLELPPSP 1352
QY 547 TKBPAPTAPE-PAPTTPKETAPTTPKKLTPTTPEKLAFTTPEKAPTTTPEELAPTTPE 605
DB 1353 RKAPKTAAPKETPAPS--PEGATTAPVQI-PPSPKSGSKKAGSKETPTTP-----SPEG 1403
QY 606 PTTTPEEP--APTTPKAAAP-----NTPKEPAPTTKBPAPTTTK 644
DB 1404 VTAAPLEIPISSKKTSMASPKETLVTPSSKKLSQTVGPKETSLEGATVPLEIPSSHKK 1463
QY 645 EPAPTTKBPAPTTKGTAPTTILKEBPAPTTKBPAPKELAP---TTTKEPTSTTSDKPA 700
DB 1464 APKVDPKQVPLTPSPKADPTLLAE-SFSPKPK-APKTAAPPSEVTVTPP-----EKPA 1516
QY 701 PTTPKGTAPTTKBPAPTTTKEPA-----PTTPKGTAPTTLKEBPAPTTPKKBPAPKE 751
DB 1517 -TPQASGTTASKVPVPAETQEVAVSSRETPTVPAVPPVKNVPSHKKTSKTIELKEAPAT 1575
QY 752 LAPTTTKGPTSTSDKBPAPTTKBPAPTTKBPAPTTKBPAPTTTETPPTTSEVSTPT 811
DB 1576 LPSPSTKSPKPSKK---APTSAP-----KEFPASISIKPVTLSLAQT-- 1617
QY 812 TTKEPTTIHKSPEDESTE--LSAEPKPALENSPKBPVPTTKTPAATKPEMTTAKDKT 869
DB 1618 --APSLQKAPSTTIPKENLAA--PAVLPSVSSKSPAAP-ARASASLSLSP--ATAAPQT 1667
QY 870 TERDLRTP-----ETTTA-----APKMTKETATTTKTESKITATTQVTS 912
DB 1668 APKEATTIPSCKAAATETPIETSTAPSLGAPKETSETSV-----SKVLSSPPKKA 1720
QY 913 TTTQDTPPKITL---KTTILAPKVTTTKITTTTEIMNKPEETAKPKDRATNSKAT 968
DB 1721 SSSKRASTLPAITLPSLKEASVLSPTATSGK-----DSHISPVSDACSTGTTT 1769
QY 969 PKQKPTKAPK-----PTSTK 985
DB 1770 --PQASEKLSKXGPTAFTTEMLAAPAPESALAITAPIQKSPGANSNSASSPKCPDPSSKK 1827
QY 986 KPCTMPR--VRKPKTTPPKMTSTWPELNPTSRIAEMLQTTTRNQ----- 1031
DB 1828 DTGCLPSAVALAPQTVPEK-----DTSKAIETLLVSPAKGSDCLHSPKGPVGSQ 1877
QY 1032 -----TPNSKLVENVPKSADGAEGETP 1055
DB 1878 VATPLAFTSDKVPPEAVSASVAPKAPAPASLTLAPSVAPLPPKQPLLESAPGVSLESP 1937
QY 1056 HMLLRHVFMPEVTPDMDYLPRV-----PNQGIILN-----P 1087
DB 1938 SKL-----PVPAEDELPPILPPEAVSGGEPFQDILVNMPPAKPAGTAPAPSAKQP 1989
QY 1088 ML 1089
DB 1990 VL 1991

RESULT 8	Db	221	STTSTSSSTTS	235
S48478	Qy	177	DNGDFKVTTPDSTTOHNVSTSPKITTAKINP	228
Glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)	Db	236	-----STSESSTTSSESSTSSSTAPATPTTSCTEKTPPTTSCTEKTPPP	288
N/Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YI019c	Qy	229	NKETTVETKETTNNKQSTDKKETTSAKETQSEKSAKDLAPTSLKVLAKPTKAPT	288
C/Species: Saccharomyces cerevisiae	Db	289	HHDTTPTCTTKKTTSK-TCT---KTTTPVPTPS--SSITSSAPV-----PTSSSTT	337
C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004	Qy	289	TKGPALTTTPE-----PTTTPKEPASTTTPKEPTTTIKSAP-----TTPKEPAPT	336
C/Accession: S48478; A26877; E26877; S27281; J06123	Db	338	ESSAPVTSSTTSSAPVPTSSSTSSAPVTSSTSSAPVTSSTSSAPVPT	397
R/Rowley, K.	Qy	337	KGAPTTTKEPAP-----TTPKEPAP-----TTPKEPAP-----TTPKEPAP	372
submitted to the EMBL Data Library, October 1994	Db	398	PSSSTTSSAPVTSSTTSSAPVTSSTTSSAPVTSSTTSSAPVTSSTTSSAP	457
A/Reference number: S48478	Qy	373	TITKGAPTTKEPAPTTPKAPATTTPKAPATTTPKEPTP---TTPKEPATTKE	423
A/Accession: S48478	Db	458	VTPSSSTTSSAPV---SSITSSAPV---PTSSSTTSSAPVTSSTTSS	509
A/Molecule type: DNA	Qy	424	---PAP---TTPKEPAPTAPKAPATTTPKAPATT---PKEPAPT---	462
A/Residues: 1-1367 <ROW>	Db	510	APVPTPSSSTTSSAP-APTSSSTTSSAPVTSSTTSSAPVPTSSSTTSS	568
A/Cross-references: UNIPROT:P08640; GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; GS	Qy	463	BSPTTKEPAPTTPKAPATTTPKAPATTTPKAPATT---PKEPAPTTPKEPAP	512
R/Yamashita, I.; Nakamura, M.; Fukui, S.	Db	569	VTSSTTSSAPVPTSSSTTSSAPVPTSSSTTSSAPVPTSSSTTSSAPVTS	628
J. Bacteriol. 169, 2142-2149, 1987	Qy	513	---TTPKEPAPTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAP	570
A/Title: Gene fusion is a possible mechanism underlying the evolution of STAL.	Db	629	SITESAPVPTSSSTTSSAPVPT---PSSSTTSSAPVPTSSSTTSSAPV	684
A/Reference number: A91831; MUID:87194600; PMID:3106330	Qy	571	PKLPTTPEKLAPTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAP	626
A/Accession: A26877	Db	685	---SSTTSSAPV---SSTTSSAPV---PTSSSTTSSAPVPTSSSTTSS	732
A/Molecule type: DNA	Qy	627	PKEPAP-----TTPKEPAPTTPKAPATTTPKAPATTTPKAPATTTPKAP	680
A/Residues: 762-1331 <YA2>	Db	733	SSAPVPTSSSTTSSAPV---SSITSSAPVPTSSSTTSSAPVPTSSSTT	788
A/Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526	Qy	681	KELAPTTPKEPTSTSDK---PAPTPKGTAPTTPKAPATT---	725
R/Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.	Db	789	SSAPVPT---PSSSTTSSAPVPT---TPSSSTTSSAPVPTSSSTTSS	845
FEBS Lett. 239, 179-184, 1988	Qy	726	TPKGTAP---TTLKEPAPTTPKAPKELAPTTPKGP---TSTSDKAPTTPK	776
A/Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar	Db	846	TTESSAPVSSSTTSSAPV---PTSSSTTSSAPVPTSSSTTSSAPVPTSS	898
A/Reference number: S27281; MUID:89031230; PMID:3141213	Qy	777	APTTPKEPAPTTPKAPATTTPET---PPPTSEVSTPTTPKEPTTIHKSPD	834
A/Accession: S27281	Db	899	TPSSSKYFGSQTETSVSSTTETIVPTKTTTSTVPTTITTTTTCSTGINS	958
A/Molecule type: DNA	Qy	835	TPKALENSPKGPVPTTKTTPATKPEMTT-----TAKDKTTERDLRTT	886
A/Residues: 1-1367 <ROW>	Db	959	SPKVTIT---VPTTITTSSTTSSITTTTTCSTGINSAGTTSCGSKTIT	1013
A/Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525	Qy	887	MTKEATTTEKTESKITATTQVST---TTQDTTPPK-----ITLKATTL	931
A/Accession: B26877	Db	1014	STSPSETASESTTTPVTTVSTTWTTEYSTTKPGGEITTTFTVKIPTT	1073
A/Molecule type: DNA	Qy	932	A---PKVTT---TKKTTITTEI---MKNPBT---AKPK-----	958
A/Residues: 1-242 <YAM>	Db	1074	APTBSVITVPTTPTTITTTTCSTGINSAGTTSGCSPKVTITTVTGSTG	1133
A/Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525	Qy	959	-----DRATNS---KATT-----PKQKPTKAPKP-----	981
A/Accession: B26877	Db	1134	LVTAVTTTAVTTTSSGTNSAGTTTGYTKSVPTTWTTLAPSAVPTPATNA	1193
A/Molecule type: DNA	Qy	982	-----TSYKKPKTM-----PRVRKPTTTPKMTSTPELN---	1013
A/Residues: 1-242 <YAM>	Db	117	SSSSSSSSSTIWKISSNSAANRELQKLKVNKNKRTKKKTPKPPVDEAG	176
A/Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526	Qy	177	DNGDFKVTTPDSTTOHNVSTSPKITTAKINP	228
R/Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.	Db	236	-----STSESSTTSSESSTSSSTAPATPTTSCTEKTPPTTSCTEKTPPP	288
FEBS Lett. 239, 179-184, 1988	Qy	229	NKETTVETKETTNNKQSTDKKETTSAKETQSEKSAKDLAPTSLKVLAKPTKAPT	288
A/Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar	Db	289	HHDTTPTCTTKKTTSK-TCT---KTTTPVPTPS--SSITSSAPV-----PTSSSTT	337
A/Reference number: S27281; MUID:89031230; PMID:3141213	Qy	289	TKGPALTTTPE-----PTTTPKEPASTTTPKEPTTTIKSAP-----TTPKEPAPT	336
A/Accession: S27281	Db	338	ESSAPVTSSTTSSAPVPTSSSTSSAPVTSSTSSAPVTSSTSSAPVPT	397
A/Molecule type: DNA	Qy	337	KGAPTTTKEPAP-----TTPKEPAP-----TTPKEPAP-----TTPKEPAP	372
A/Residues: 1-1367 <ROW>	Db	398	PSSSTTSSAPVTSSTTSSAPVTSSTTSSAPVTSSTTSSAPVTSSTTSSAP	457
A/Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525	Qy	373	TITKGAPTTKEPAPTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAP	423
A/Accession: B26877	Db	458	VTPSSSTTSSAPV---SSITSSAPV---PTSSSTTSSAPVTSSTTSS	509
A/Molecule type: DNA	Qy	424	---PAP---TTPKEPAPTAPKAPATTTPKAPATT---PKEPAPT---	462
A/Residues: 762-1331 <YA2>	Db	510	APVPTPSSSTTSSAP-APTSSSTTSSAPVTSSTTSSAPVPTSSSTTSS	568
A/Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526	Qy	463	BSPTTKEPAPTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAP	512
R/Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.	Db	569	VTSSTTSSAPVPTSSSTTSSAPVPTSSSTTSSAPVPTSSSTTSSAPVTS	628
FEBS Lett. 239, 179-184, 1988	Qy	513	---TTPKEPAPTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAP	570
A/Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar	Db	629	SITESAPVPTSSSTTSSAPVPT---PSSSTTSSAPVPTSSSTTSSAPV	684
A/Reference number: S27281; MUID:89031230; PMID:3141213	Qy	571	PKLPTTPEKLAPTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAP	626
A/Accession: S27281	Db	685	---SSTTSSAPV---SSTTSSAPV---PTSSSTTSSAPVPTSSSTTSS	732
A/Molecule type: DNA	Qy	627	PKEPAP-----TTPKEPAPTTPKAPATTTPKAPATTTPKAPATTTPKAP	680
A/Residues: 1-1367 <ROW>	Db	733	SSAPVPTSSSTTSSAPV---SSITSSAPVPTSSSTTSSAPVPTSSSTT	788
A/Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525	Qy	681	KELAPTTPKEPTSTSDK---PAPTPKGTAPTTPKAPATT---	725
A/Accession: B26877	Db	789	SSAPVPT---PSSSTTSSAPVPT---TPSSSTTSSAPVPTSSSTTSS	845
A/Molecule type: DNA	Qy	726	TPKGTAP---TTLKEPAPTTPKAPKELAPTTPKGP---TSTSDKAPTTPK	776
A/Residues: 1-1367 <ROW>	Db	846	TTESSAPVSSSTTSSAPV---PTSSSTTSSAPVPTSSSTTSSAPVPTSS	898
A/Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525	Qy	777	APTTPKEPAPTTPKAPATTTPET---PPPTSEVSTPTTPKEPTTIHKSPD	834
A/Accession: B26877	Db	899	TPSSSKYFGSQTETSVSSTTETIVPTKTTTSTVPTTITTTTTCSTGINS	958
A/Molecule type: DNA	Qy	835	TPKALENSPKGPVPTTKTTPATKPEMTT-----TAKDKTTERDLRTT	886
A/Residues: 1-1367 <ROW>	Db	959	SPKVTIT---VPTTITTSSTTSSITTTTTCSTGINSAGTTSCGSKTIT	1013
A/Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525	Qy	887	MTKEATTTEKTESKITATTQVST---TTQDTTPPK-----ITLKATTL	931
A/Accession: B26877	Db	1014	STSPSETASESTTTPVTTVSTTWTTEYSTTKPGGEITTTFTVKIPTT	1073
A/Molecule type: DNA	Qy	932	A---PKVTT---TKKTTITTEI---MKNPBT---AKPK-----	958
A/Residues: 1-242 <YAM>	Db	1074	APTBSVITVPTTPTTITTTTCSTGINSAGTTSGCSPKVTITTVTGSTG	1133
A/Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525	Qy	959	-----DRATNS---KATT-----PKQKPTKAPKP-----	981
A/Accession: B26877	Db	1134	LVTAVTTTAVTTTSSGTNSAGTTTGYTKSVPTTWTTLAPSAVPTPATNA	1193
A/Molecule type: DNA	Qy	982	-----TSYKKPKTM-----PRVRKPTTTPKMTSTPELN---	1013

Query Match 15.9%; Score 928.3; DB 1; Length 1367;
 Best Local Similarity 23.8%; Pred. No. 1.9e-11;
 Matches 347; Conservative 110; Mismatches 456; Indels 547; Gaps 65;
 13 LLVFEVIQVSSQDLSACGACGSGYSDATC-----NCDYN-----49
 6 LLVVLVLSLNSALGFPTALVPRGSEGTSCNSVNGCPNLDNFNMHMOQNTMQVTLGV 65
 50 -----49
 66 TSVSWQDNTYQITIHVKGENIDKYLWSLKIIGVTGPKGTQVLYGNENTYLDNPTD 125
 50 -----76
 126 FTATFEVATQDVNSQVWM---PNFQI-----FEYQSGAAQYASGWGTT 171
 77 SFERGEC-----DCDAQCKYDKCCPDYBSFCABEHSVSNQES 116
 172 SFDLSTGCNNYDNGHQSOTDPFGYWNIDCDNNC-----GGTKSSTTSSSES 220
 117 SSSSSSSSSSTIWKISSNSAANRELQKLKVNKNKRTKKKTPKPPVDEAGSGL 176

Db 1194 TTECSAATNAAGETTIVCSAKTIVSASAGENTAPSATTPVTTAIPTTVITTESSVGINS 1253
 QY 1014 -----PSSIAEAM-----LQTTTPNQTPNSKLVENVNPKSEDAGG 1049
 Db 1254 AGETTTGTTKSIPIYITLIPGNGAKNYETATNP-----ISIKTTSQLAT 1305
 QY 1050 AGETPHMLLRPHVFMPEVT 1069
 Db 1306 AGASS---VAPVWTSPLT 1321
 RESULT 9
 T34513
 hypothetical protein ZK783.1 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
 C;Accession: T34513
 R;Favella, A.; Vaudin, M.
 submitted to the EMBL Data Library, August 1994
 A;Description: The sequence of C. elegans cosmid ZK783.
 A;Reference number: Z21536
 A;Accession: T34513
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 13507 <FAV>
 A;Cross-references: UNIPROT:Q23587; EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN000021; CESP:ZX
 A;Experimental source: strain Bristol N2; clone ZK783
 C;Genetics:
 A;Gene: CESP:ZK783.1
 A;Map position: 3
 A;introns: 14/1, 48/2, 84/1, 196/3, 303/1, 381/1, 586/1, 605/1, 1175/3, 1207/1, 1409/2, 3504/1
 Query Match 15.5%; Score 901.5; DB 2; Length 3507;
 Best Local Similarity 13.0%; Pred. No. 2.3e-10;
 Matches 396; Conservative 186; Mismatches 450; Indels 2005; Gaps 90;
 QY 7 PIVLLLLSVFVITQYSS-----QDL-----SSC 30
 Db 32 PTFVNFDSTVICQSSDPTDLHINMSSLCDGKQDCFNVPAMHDEVFPYCEHKQSTC 91
 QY 31 AGR-----CGEYSRDA-----TCNC----- 46
 Db 92 SGKGACLYDGAKPCYCDGSGSACELQDKBLEHPCHPMMAQCQTLGSEYECRLPGY 151
 QY 47 ---DYNCOHMEC-----CPDFKV----- 63
 Db 152 EGNHETDIDECDKLTSCPEHSCINLPGTYCNCTQGFPPKNGQSGLDKCADINE 211
 QY 64 -----CTAELCK-----GRCPESFERGRE----- 83
 Db 212 CETGAHNCDADEICENSIGSFKNKSPGYELIDGKCEDVNECGSEKLHKCDVRACVN 271
 QY 84 -----DAOCK-----CDC-- 86
 Db 272 TIGYECEBEGEGDGKQKQKCGHVKLPFSKSCRKNSAICDRHASCHIVLDICDCKT 331
 QY 87 -----DAOCK----- 92
 Db 332 GYTGDIATCHINECDAKOTPCSDGGRCLNLDGYVCCCKNGQDDATCICKQAFCSGGCG 391
 QY 93 -----YDKC-----CPDYESFAE-----EHSVSE 112
 Db 392 DNAICSNATCACIDGFRGDPHKKVDINECVENDSVCGGVGDCVNLFGGFKCCQHGSTE 451
 QY 113 -----NQESSSSSSSS----- 125
 Db 452 ABCTDQAFSSDSTISSHGADFTTGEQIIEGSGSIQTSSGGSITVTRGLIPKDVLTTS 511
 QY 126 ----- 125
 Db 512 GRLACTSYCPNPSECVGCEYGVGGNALVGCEDIDECITETICNIEANWCVNLIIGF 571

QY 126 ----- 125
 Db 572 VCCNPTNATHDDCIDFLTKVIVAYMIIIFLLKGLKLEITKEKGLHVIIGNBEDTVVATRSN 631
 QY 126 ----- 125
 Db 632 HSTDQLITQVVQSRNFTSQIILTRKVSSEAVTQTTDADEFGLEISAADLAGSGSG 691
 QY 126 -----SSTW----- 130
 Db 692 ITLPTLEPKIEGSKKASGGVWTEDEDEDLMEEGSGSWSTTINGTIGTSPRSEGT 751
 QY 131 -----KIKSKNSAANRELQKKLV-KDNK----- 154
 Db 752 IRRVITTLGEDGEPETATKPCISAPDKTGEKSTESDGEKLTVEKDGAQAQSGSGSATS 811
 QY 155 -----KURTKKTPKPPVVDAGSGL-----DNGDFKVT-- 185
 Db 812 SGKSEATSGSSSSAKSGTGSEASGSSGSSGSGVSGSGSVSTESGFTSSSGS 871
 QY 186 -----PDTSTTOHNVST-----SPKITTAKPIN 209
 Db 872 VGSEATGTGVDGSESGKPSKSTTEKLPTKNGEKSPISGSDTTGKESSEETTSRKPIE 931
 QY 210 PRPSLPNNS-----DTSKETSITLVNK-----ET 232
 Db 932 GSDSLTEGSGSEWFEETGSKGHFESGKSVTSCKGTQSGAEGSGSGPKVPKGPAPAI 991
 QY 233 TVETKETT-----NKQSTDGKEKTSKAKETOSIEKTS 267
 Db 992 TTDGESSSTGDKSGKPADKSDKNVPKTDGKNDPDIITDGEDSTSETSGGQPKGK 1051
 QY 268 AKDLAP-----TSKVLAKPT----- 282
 Db 1052 SKGPPGDKGEVKKPTEVDGPNLGGTKGKSNVPLKPTDLPEEGSGLTSSGGKNST 1111
 QY 283 -----PK-----AETTKGPAITPKETPT--TPKE-----PASTT- 312
 Db 1112 FEHGKLERLPKTKEDSKSETPOLGLEISACKPEPEDGTSKEVGLLEILWESTTPGSTTL 1171
 QY 313 ----- 312
 Db 1172 DSDSVGLEISGSDITKATKKPHEIGSGTGDEBITATRDVSKSTKKRVEVDGNGE 1231
 QY 313 -----PKEPTPTIKSA-----PTT-----PKP----- 331
 Db 1232 TSGVDGKTTTAPTPSSAESSTSRIPPTTSEASPEGSGGAGVPESPDGSGESSTSPDG 1291
 QY 332 -APTTKSAPTTPKEPAPT-----TKEPAPTTPKPAPT----- 366
 Db 1292 VSPSTATAPEVPTTSASSTPDAVEESGIPSTSKPTAPLETT-----APSTEVSPEGSG 1347
 QY 367 -----TKEPAPTTKSAPTTPKEPAPT-----KKPAPTTPKPAPT----- 405
 Db 1348 TEESTLPTEGSGESTTSSAPT-----EPATVLPQNRNEKPEPTKDTFALPTTTGAPAN 1405
 QY 406 -----PK----- 408
 Db 1406 DSSVENTKCTSSDEGLDALCERRTGVCRBPGFEGAPPKKSCVDVDECATGDHCHESA 1465
 QY 409 -----PT----- 410
 Db 1466 RCQNVYGVYACFPCTGFRKADGSGQDIDECTEHNSTCCGANAKCVNPKGTYSECENG 1525
 QY 411 -----PTTPKEPAPTTK----- 422
 Db 1526 LDGQYCCVPTT-KKPCDSTQSSKSHCSBSNWSCEVDTVDGSVECKECMGYKSGKVCED 1584
 QY 423 -----BPAP----- 426
 Db 1585 INECVABKAPCSLNANCMNMGTFSCCKQYRGDGFMTDINECDERHPCHPHAECTNL 1644
 QY 427 -----TPKE----- 431

Db 1645 EGSFKCEHSGFEGDGKIKCTNPLERSCEBVEKFCGRVDHVHSLSVIYNGSLSSVCECE 1704
QY 432 -----PA----- 433
Db 1705 PGRFPEKESNCVDIDECEBSRNNCDPASAVCVNTEGSRCECAGYEGEGVCTDIDEC 1764
QY 434 -----PTAPK----- 438
Db 1765 DRGAGCDNMAGINRMGSCGCKMAGYTGCGATCIKIBEEPCKDKTACTDEWSLCELE 1824
QY 439 KPAPTTPKSPAP----- 450
Db 1825 KKQCTVDESEVPCGACLGHPHNGTCOSLOISGLCAQKNDNCNKAECIDIHPDSHFCS 1884
QY 451 -----TPKE 455
Db 1885 CPDGFIDGMI CDDVDECNAGCDDNTKCENTTIGSFNCVLEGFKKVDKCVVDEKKQ 1944
QY 456 P-----APTTKEPSPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTT 503
Db 1945 PNRKEIIDEENSSNSGQKPTTKGIVSSTSATSSSESTTAEPHVTTISSTSTKDMT 2004
QY 504 PTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP 563
Db 2005 SSKSPENVMTSSSEPSVSSSKSTASSTVSSSPSS-----SSAPLTSSPATT 2057
QY 564 KETAPTPKLTPTPE-----KLAPTP-----KPAPTTPELAPTPPEPTP 608
Db 2058 TEV-----ITESSVKSTTPKESSEITVKLSSKSPETVSSKSPSTPS-----TTSQSVTS 2111
QY 609 TPEEPAPTPPKAAAPNTPKP-----APTPKEPAPTTTPKEPAPTTPKET-----AP 656
Db 2112 TVPETSIVLSEAPVTSSTEVHTSETKPSLSASSTGTDNSTTPSSSLASVKST 2171
QY 657 TTPKGT-----APTTLKEPAPTTTPKAPKELAPTT-----TTKEP----- 691
Db 2172 SAPEGTSASVAPVLSLSSPDV-SQPSTKTFATESSTVQASETSSGTSVKSTSEPSHV 2230
QY 692 --TSTTSDKAPTPTPKGAAPTPKAPTPKAPTPKAPTPKAPTPKAPTPKAPTPKAP 730
Db 2231 TKLSITSSNPSSVPVTPKSTPTTPEST--EQPTSTTPSQSLTPMNSNSELVTTSEPH 2288
QY 731 -----TAPTLKEPAPT-TPKFPAPKEL---APTTPKGT-----STT- 764
Db 2289 VLSSLSLSPDVQSSTTPNNLSSESTVETPKTSSEVSLNSEPSTTEAPTTLSPDILSTT 2348
QY 765 -----SDKP-----APTTP----- 773
Db 2349 NNLQSOSTVSTEDREISENSESEKPTSAPELVTSSTVTHVASSPDVPTSESPDPLTGSS 2408
QY 774 -----KETAPTP-----KEPAPTTPKKP-----APTTP----- 798
Db 2409 TENIPEASSKQIISPTTPDITTAEEPTKSTNSPDLSITTSNVLSSESTTPESSSKSPV 2468
QY 799 -----TPPTTSEVSTPTTPKEPTT----- 818
Db 2469 SSSTEGISVVTSTEFKVPKESTISSVLLEDLTTPSPILEETTTASETSEPLTSLTV 2528
QY 819 ---IHK--SPDBSTPELGAETPKALENSP-KEPG-----VPTT----- 851
Db 2529 SVRIHELITSSENVKESSEITTSSESKPQEPAGILTSVVVPTSSVLITASEIEAI 2588
QY 852 -----KTPAATKPEM-----TTAKDKTTERDLTTPETTTAAKPKTKETATT 894
Db 2589 TSNTPFKQGRTPITTPSPKSLVKSTTSPSTVTSSEPSSESTKRTTSTVSTTTPTEETTS 2648
QY 895 -----TEXTTSKITATTQVTS-----TTQDTTTPFKI 923
Db 2649 ESLIITAAPSKPTSESTTSSSEAPTTPAKTSETKPSNVSTGRKSTENVETSTSGSLES 2708
QY 924 TLLKTT-----TLAPKVTTTKTIPTT----- 945

Db 2709 STMSTSSSEPTNAPAVTVSSSEASTTLENSSTSTSEASTSKLSSLPFSEITSEAVT 2768
QY 946 -----EIMNKPEE----- 953
Db 2769 VSSRAPAEITWSSSHREISTVSSSESEPEIPLSTTVSPNVVTASSIPSEEPILSSVTSS 2828
QY 954 -----TAKPKD-----RATNSKA-----TTPKPKOKPT 975
Db 2829 STPRVRLITGTPDDLIVSVTVPSHGRRQNITASSVPSNSTSPIILPSESLLTPQPPPTT 2888
QY 976 KAPKPKSTTKPKTMRVRKPK---TTPTP-----RKMTSTMPELNPTS 1016
Db 2889 TTTAKPATTSGKGPSPSIQPPAEMFTTPAPPPPSNGYGEETNQEEQVST-----TT 2942
QY 1017 RIAEAMLOQ-----TTRNQTPNSKLVEVNPKSEDAAG 1050
Db 2943 TEAPSLCSTVTCHSLATCEQSTGVCICRDGFIGDTTACSKKSTADICISLPSLCADKAC 3002
QY 1051 EGET-----PH---MLLRPHVFMPE 1067
Db 3003 DNSTRCECDAGVIGDGYVCSPPHPCDVLRLDNLCSPE 3039
RESULT 10
S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 09-Jul-2004
C:Accession: S49915
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A:Description: Pex genes: pollen-specific genes with extensin-like domains.
A:Reference number: S49915
A:Accession: S49915
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1188 <RUB>
A:Cross-references: UNIPROT:Q41805; EMBL:Z34465; NID:g600117; PID:g60011
Query Match 15.3%; Score 888.9; DB 2; Length 1188;
Best Local Similarity 20.1%; Pred. No. 9.2e-11;
Matches 289; Conservative 87; Mismatches 423; Indels 641; Gaps 54;
QY 10 LLLL--LSVFTVQVSSQDLSSCA-----GRCGE----- 36
Db 17 LLLLAACLSACSQAVTSAEASVIAHQLLAMKEAGGEGADLPADFEFDDRVGAANFN 76
QY 37 -----GYSRDATA-----CDYNQHYMECCPDFKRVCTAEL 68
Db 77 PLRRAYIALQAWHRAFYSDPKGY---TANWGEDVCKYNGVICTEALDDPKITVAGI 132
QY 69 SCKGR----- 73
Db 133 DLNGADIAGVLPPELGLLTDLAFFHINTNRFPGIIPKMSRLSLLHEPDVSNRFFVGVP 192
QY 74 --CFES-----FERGECDDA----- 88
Db 193 YVCEMLSVLKYLDLRFNDFEGELPALEFK---DLDAIFVNTNRFVGPENLGNSTAS 248
QY 89 -----QC-----KKYDKCCP-----DYESFCAEE 107
Db 249 VIVFANNAFVGIPKISGRMVKTLDEIIFLNKLQCLPLEMGLLVNTVID----- 300
QY 108 HVSVENQESSSSSSSSSSSTTWIKSSKN-----S 138
Db 301 --VSGNMLVTGPIEQULSIAKLEQLDVSERNVFTGVHESICELPALVNFSAFNFNSEA 358
QY 139 AANRELQKLVKVDKNKN---RTKKK-----PTKPEPVVD 170
Db 359 AVCMPSDKALVNLDDRDNCLGALRPAQKTALOCAPVLRPVDCSHVCAGVPTP----- 412
QY 171 EAGSGLDNGDFKVTTPDTSTTOHNVKSTSPKITTAKPINRPSLPSPNSDTSKETSLTVNK 230

Db 610 STTEB--PTTDEPTTTT---ESSTGKATPBLSTTSEBTTELK---ITTE----- 655
 QY 682 ELAPTTTKEPTSTSDKAP---TTPKGTAPTTKPEAPTTKPEAPTTKPGTAPTTLK 737
 Db 656 --GSTTTEPTTALFAEASGIIITDBETTTSTTSTTPEITSTKE--IVTBSA:QTQSVS 711
 QY 738 BPAPTP---KXAPKELAPTTTKGP----- 760
 Db 712 VVESSTPQLPBRKAIUNVKHNLVLEKXRLKKEKESSTTSGSSSETTTVAENID 771
 QY 761 --TSTSDKAPTTKETAPTTKPEAPTTKPKAPTTPE--TPPTTSE-VSTPTTKE 815
 Db 772 EVTTTEKEKVQTTPTTTEKSTQSETTTTTTTTEKTTSTKTTTEKPTTSASATTETTSE 831
 QY 816 PTTIHKSDPESTPELSAPTPKALENSPKPGVPTTKTPAATKPEMTTAKDKTTERDLR 875
 Db 832 PST-----TEST-----TVDTSATTBESSTAETTTTISAE-- 862
 QY 876 TTPETTTA-----APKMTKETATTTE 896
 Db 863 -TSETTTSESAAFITGESPENTALOSSQSKSEENESSAEKPGARRDFVPKHKTTVKPAE 921
 QY 897 KTTESKITAATTO-VTST--TTQDTPPKITTLKTTTLAPKV7---TTKKTITTTIMN 949
 Db 922 TTSVAASATTTTETPTTTEKSTTLETTTLETTTLETTTLETTTLETTTLETTTLETT 981
 QY 950 KPEETAKPKDRATNSKATTPKPO-----KPTKAPKPTSTKPKTMPR 992
 Db 982 K-----INNTQISQPKPTDISKTALSSLLISGLIGSFTKAPMAPTI----- 1022
 QY 993 VAKPKTTTPPKMTSTMPLN-----PTSRIAEAMLOTTPRNPNTPKLVEV 1040
 Db 1023 ---HTTTDAAPVTAATEASLNDGSKIIDEAQPTDEIRRA-----OPTN----- 1063
 QY 1041 NPKSEDAGGAGETHTHMLLRPHVFMPEVTPMDVLPVPYNGI 1083
 Db 1064 -----EMDKEMEFKRIREQRI 1080

RESULT 12
 T29757
 Protein UNC-89 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
 C:Accession: T29757
 R:Du, Z.; Le, T.T.; Wilson, R.
 submitted to the EMBL Data Library, May 1997
 A:Description: The sequence of C. elegans cosmid C09D1.
 A:Reference number: Z20679
 A:Accession: T29757
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-6642 <DUZ>
 A:Cross-references: EMBL:AF003131; PIDN:AA854132.1; GSPDB:GN00019; CESP:unc-89
 A:Experimental source: strain Bristol N2; clone C09D1
 C:Genetics:
 A:Gene: CESP:unc-89
 A:Map position: 1
 A:Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6
 /3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Query Match 14.8%; Score 862.5; DB 2; Length 6642;
 Best Local Similarity 13.9%; Pred. No. 3.3e-09;
 Matches 397; Conservative 146; Mismatches 385; Indels 1925; Gaps 96;
 34 CEGYSRDATCNC-----DYNCOH----- 52
 Db 659 CKEGDVDFCEVEGWPPPELWLVDDQPLRPSHDFRLQYDQOTAKLEIRDAQPDGTGY 718
 QY 53 -----YMECCPD-----FKRYCTAE-- 67
 Db 719 TVKIQNEGSGISBKAELFVQADPDKNHVAPEQATIEVEEC--DEGEVRFKSVITGPN 776

QY 68 -----LSCKGRCFESPERGECDC 86
 Db 777 PELIWFINKPLSESEKVKFISEDGICILITIKDVTTRHFDGMVTCGS-----NRLGSASC 831
 QY 87 DAQCK-----KYDKCCPD-----YE-----SFAAE----- 107
 Db 832 DGRKLVKRVPPAPPTFNKPLEDKTVQEKSTVVPEDVSGWPEPTLTFLTGLCKELKNGEGV 891
 QY 108 -----HSYSEN-----QESSSSSS----- 122
 Db 892 EIVGHDGFYRISIPNTSMKDHGEIVAKAQNDEHGTAEGRALUTVQEESRSAPTFLKD 951
 QY 123 -----SSSSSTIWKIKSK----- 136
 Db 952 IEDQTVKTGEFAVFETTVRGNNPENVTFWINGHKMDQSGPGVKIEAHNHDHKLITIDSAQY 1011
 QY 137 -----NSAANRELQKLVKONKKTURTAKKP----- 162
 Db 1012 AGTVLCRAENAVGRFETKARLVLAPEK--OKKPKFVFEILVDKTETVDNTVFEVAVE 1068
 QY 163 -TPKPPV-----VDEAG-----SGLDNGD 180
 Db 1069 GEPKPTVWYLGEBELKQSDRVEIREFDGSIKISIKNIKIEDAGELRAVATNSESDET 1128
 QY 181 FKVT----- 184
 Db 1129 AKLTVOQKPPFAPEFDLRPVSLTVEKGEAVFSAHAFGLPTTYEWSVNGRKYRQDQEGAR 1188
 QY 185 -----TPDTST-----TQHNKVS----- 197
 Db 1189 VTRDESTVDCASILLIDTAYYSEVNNHLTISVVAENTLGAETGAQLIEPKESVWVEK 1248
 QY 198 -----TSPKITTAKPINRPSLPNSDTSKETSLSLTWNKETTETKET 239
 Db 1249 QDLSSSEVQKIAQVKEASPEATTI-----ITMETSLSITKTTTMSITTEV 1294
 QY 240 TTT----- 242
 Db 1295 TSTVGGVTVEKSESESATTVIGGGGVVTEGSIKIEVSVKTDQTDVREGTPRR 1354
 QY 243 -----NKQTSIDGKEKTTSAKETOSIEKTSARDLAPTSKVLAKPTP 283
 Db 1355 VSPABEELPKEVIDSDRKKKSPSPDKKEK-----SPEKTEKPSPTKKT----- 1400
 QY 284 KAETTTKGALTTKPEPTTTP---KEPASTTPKEPT-----PTIKSAPTTP--- 328
 Db 1401 -----GEEVKSPEKSPASPTKKEKSPAADEVKSPTKKEKSPSPPTKKEKSPSPPTKK 1453
 QY 329 -----KEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTP---KEPAPTTT--- 367
 Db 1454 TGDEVKEKSP-----PKSP---TKKEKSPKEDVKSPVKKEKSPDANIVEVSSE 1501
 QY 368 -----KEPAPTTTKSAPTTP----- 382
 Db 1502 TTIEKTETTMTTMTHESESRTSVKKEKTPKDEKPKSPTKKDKSPKESITESIKSPV 1561
 QY 383 -KEPAP-TTPKPKAPTTKPEAPTTKPEPTTTPKE-PAPTTKPEAP-----TTPKE 431
 Db 1562 KKEKSPKVEKPEKSPKKEKSPKASPTKSENEVKSPTKKEKSPKESVVEELKSPKE 1621
 QY 432 PAP-TAPKPKAPTTKPEAPTTKPEAPTTTKEPAPTTTKEPAPTTTTSAPTTPKEAPT 490
 Db 1622 KSPEKADDPKSPKKE-----KSPKESATSDVKSPKKEKSPKVEKPT-----SPT 1670
 QY 491 TTKSAPT-----TPKEPSPTTKE-PAPTTKPEAP-----TTPKPKAPTTP 531
 Db 1671 KKESSPTKTTDDDEVKSPTKKEKSPQVEEKASPTKKEKSPKESVVEVKSPEKSPKEA 1730
 QY 532 KB-PAPTTPKEAP-----TTTKK---PAPTAPKEPAPTTPKETAPT----- 569
 Db 1731 EEPKPSPTKKEKSPKESAAAEVKSPTKKEKSPKESAAEEKPSPTKKEKSPVQADDEVKS 1790
 QY 570 -----TPKKL-----TPTTPEKLAPTTPKEPAPTTPPELAPTTPPEPTPTTP----- 611

Db 1791 PTKKSPKVEKSPSPKKEK-----TPKSA---AELKSPTKKESPPSPKTKTGDE 1843
QY 612 -----EPAPPTPKAAAPNPKBPAPPTPKBPAPPTPKBPAP----- 648
Db 1844 SKEKSPKPEKPKSPTPKSPKSPKPK-----KKSKSPKAPKPKPAPKLTDLKLQTVNK 1898
QY 649 ----- 648
Db 1899 TDLAHFVVVHATECKWFLDGEKITTAGVTVSKDDQPEPRCSIDTTFMFGSGTVSVVAS 1958
QY 649 -----TTPKET----- 654
Db 1959 NAAGSVETKELKVLKLETPKTKPKBFTDKLRDMVETKGDVQVMDVIALHSLPYKQYQNGN 2018
QY 655 -----APTPEKGTAPPTL----- 667
Db 2019 LLEDGKNGVTIKNENKSLIIPNAQDSGKIIVASNEVSGSSSAQLTVNPPSTPIVV 2078
QY 668 KEPAPPTPKKAPKAPKELAPTTTKEPTSTTSKDPAPT----- 703
Db 2079 DGPKSVTIKE-----TETAEFKATISGFPAPTVMKWTINEKIVEESRTITTIKTEDVY 2130
QY 704 -----PKGTAPTTPK-----BPAPTTPK----- 721
Db 2131 TLKISNAKIEQGTGVKVTAGNSAGDSKQADLKVEPNVKAPKPKSQLTDKVADEGEPLRW 2190
QY 722 -----EPAPTTP-----KGTAPTTLKEPAP----- 741
Db 2191 NLELDGSPGCTEWSLLNGOPLTKSDTVQVVDHGDGTVHTIAEAKFEMSGTLTAKAKNA 2250
QY 742 -----TPPK-----KPAPK----- 750
Db 2251 AGCETSATKYTVNGGNKPEFVQAPQNHETLEESVKFSAIVTGKPMNVTVYLNNKKLI 2310
QY 751 -----ELAPTT----- 756
Db 2311 QSEEVKVKVHETGKTSIRIQRLMEHNGTIRVEAENVSGKVQATAQLKVKDKKTEVPRET 2370
QY 757 T-----KGPTSTTSDKP-----APT 772
Db 2371 TNMDRQKGEDEVKFTANVGVPEPSVAWTLNGEPVSKHPNITVTDKDGHEHTISAVT 2430
QY 773 PKET-----APTTPK----- 782
Db 2431 PEOAGELSCBATNPGSKKRDVQLAVKVKGDAPTFANLEDRLETEGETLMDAKLNIYK 2490
QY 783 ----- 782
Db 2491 PKKTIWLKGVETSDGHVKIVEEDGSLKLSILQTKLEDKGRITIKAESEFVGAECSA 2550
QY 783 -----BPAPT----- 788
Db 2551 SLGWVGRPMKAPAFQSDIAPINLTGDTLECKLLITGPTPFVKWYIGTQLVCATEDTE 2610
QY 789 -----PKK-----PAPTTPET----- 799
Db 2611 ISNANGVYTMKIHGVTADMTGKIKVAYNKAGEVSTEGPLKVAPIPVEFEISLCDATCR 2670
QY 800 -----PPPTT-----SEVSTPTTKEPT----- 818
Db 2671 EGDTLKRAVLGPEPEPWSWYNGKLEBSONIKIHSEKGTVTVTIKDITCDYSGQVVC 2730
QY 819 -----IHK----- 821
Db 2731 EAINEXKATSEATLLVLRGEPDFLEWLSNVRARTGTQVVKVVFVTDGPKESLTWYIN 2790
QY 822 -----SPD----- 824
Db 2791 NKEILNSDLTYITVDDKTSLTINSFNPVHVGEIICKAENDAGEVCTANMITYTDMF 2850
QY 825 ESTPELSAE-----PTKALENSPK-----E 845

Db 2851 SESESEAAQAEFFVGDLDLDESLREEMHRTPTPM---APKFITIKDKTKAKKHSAVFE 2907
QY 846 PGVPTTK----- 852
Db 2908 CVVPDTKGVCKWLKDGKEIELIARIVQTRTQPEGHIQTQELVLDNVNTPEDAGKYTCIVE 2967
QY 853 -----TPAAT-----KPEMTTTAKDKTTTDRDLRTT-----PETTT 882
Db 2968 NTAGKDTCEATLTVIESLEKSEKKAPEFIVALQDKTKTKSEKVLCKVIGEPKPKVSW 3027
QY 883 AAPKWTETATTTKT-TESKITATTQO---VTSTTTQDTPPKITTL---KTTTLAP 933
Db 3028 LHDNVSEKPNSEKITITQESIIVESVEGVERVTISSE-----LSHQGYTCIAE 3078
QY 934 KVTITTKT-----ITTEIMNKP-----BETAK 956
Db 3079 NTEGTSKTEAFLTVQGEAPVFTKELQNKELSIGELKVLSCSVKSGPOPHVDFYSFSETTK 3138
QY 957 PKDR-----ATNSKATTPKQKPT---KA 977
Db 3139 VETKITSSSRIAIEHDQTNTHRMVISOITKEDIVSKAIATNSIGTATSTSKITTKVEA 3198
QY 978 P-----KKPTSTK-----KP-----KTMPR----- 992
Db 3199 PVFEQGLKTSVKEKEIKMEVKVGSAPDVENFDDKPVSEDNHEMKKNPETGVFTLV 3258
QY 993 VRKPKTT-----PT-PRKMTSTWPELNPSTRIAEA 1021
Db 3259 VKQAATTDAGKYTAKASNPAAGTAESSAAEVQTSLEKPTFVRELVTTEVKINET---A 3313
QY 1022 MLQTTTR-----PNOTPNSKLV-----EV 1040
Db 3314 TLSVTVKGVPPSVEMWKDQGPVQTDSSHVIKAVEGSGSYSIITIKARLEDSGKVACRAT 3373
QY 1041 NPKSEDAAGAGETPHMLLR---PHVFMPEVTP 1070
Db 3374 NP-----AGEAKTEANFAVVKLVPPPEFVKLSP 3402
RESULT 13
B48666
Cell proliferation antigen Ki-67, short form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 15-Mar-2004
C:Accession: B48666
R:Schluster, C.; Duchrow, M.; Wohlerberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq
ins.
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: B48666
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2897 <SCH>
A:Cross-references: ENBL:X65551
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F:29-91/Domain: kinase interaction domain homology <KIH>
Query Match 14.4%; Score 839.5; DB 2; Length 2897;
Best Local Similarity 15.5%; Pred No. 3e-09;
Matches 408; Conservative 140; Mismatches 445; Indels 1635; Gaps 99;
QY 4 KTLPTYLLLLSVFVIQVSSQDLSACAGRCGEGYRDATC----- 44
Db 241 KTAPE-----ASSKSGQTEVPKRGGE---RVATCLQKRVISRSQHDILOMIC 284
QY 45 -----N 45
Db 285 SKRRSGASEANLIVAKSWADVVKLGAQTQTVIKHGFORSNKQRPAATPKPKPVGVH 344
QY 46 CDYNQOHVNECCP-----DFKR----- 62
Db 345 SQFSTGHANSPTIIIGKAHTEKHVHPARVYVRLNFI-SNQKMDFKEDLSGIAEMFKTPV 404


```

QY 555 PKPAP---TTPKET---APTTP-----KKLTPTT----- 578
Db 2103 CKSPPPESMDTPTSTRRPFKPLGKRDI VEELSALKQLTQTHDKVPGDEKGINVRE 2162
QY 579 --PEKLAP---TTPKPAPTTP-----EELA-----PTTPEETPTTPE 612
Db 2163 TAKQKLDPAASVTGSKRQPTPKGAQPLEDLAGLKELFQPVCTDPTTHEKTKIACR 2222
QY 613 EPAP-----TTPKAA-----TTPKAA----- 623
Db 2223 SPQDPVGTPTIFKPSKRSRLKADVEESLALRKRTPSVGKAMDTPKAGGDEKDKAF 2282
QY 624 -----PNTKPEPA-----PTTKEPAPTTP 643
Db 2283 MGTVPQKLDLGNLPGSKRWPTPKKAQALDLAGFKELFQTPGTDKPTTDEKTKIAC 2342
QY 644 KEPAP---TTPKET-----APTTPK----- 660
Db 2343 KSPQDPVDTPASTKQRPKNLRKADVEEFLALRKRTPSAGKAMDTPKAVSDEKINT 2402
QY 661 -----GTAPTLKEPAPTTPKPA-----KELAPTTPKPTSTSDK-- 698
Db 2403 FVETPVQKLDLGNLPGSKRO--PQTPKKALEALDLVGFKEFLQTPGCHTEESMTDKIT 2460
QY 699 -----PAP-----TTPKGTAFT----- 710
Db 2461 EVSCKSPQSEFKTSRSKQBLKPLVKVDMKEBPLAVSKLRTSGTTQTHTEPTGDSK 2520
QY 711 -----TPK--BPAPT-----TPKEPA----- 724
Db 2521 SIKAFKESPKQILDPAASVTGSRRLRTRKEKARALEDLVDFKELPSAPGHTESMTIDK 2580
QY 725 -----PTTPKGTAFTILKEAPITPKPAKELAP-----TTTKGPTSTSDKPA-- 769
Db 2581 NTKIPCKSPPELTDATSTKRCB--KTRPRKVEKEELSAVERLQTSQGTHHKFPASG 2639
QY 770 -----PTTPKE-----TAP 778
Db 2640 DEGIKVLQRAKKENPVEEBSPRRPRAPKEKAQPLEDLAGFTLSETSGHTQESITAG 2699
QY 779 TTPKEPAPTTPKPAITPETPTPTTSEVSTPTTKPTTHKSPDBSTELSAEPTP-- 836
Db 2700 KATKIPCESPLEVVDTTASTKRLHRTVQKVQKPEPSAV--RPTQTSGETTDADKEPAG 2758
QY 837 -----KALENSPKPQGVPTT-----KTPAA--TKPEMTTT 864
Db 2759 EDKGIKALKESAKOTPAASVTSRPRAPRESAQAIEDLAGFKDPAAGHTESMT-- 2816
QY 865 AKDKTTERDLTTFE---TTTAAPKMTKERTATTEKTE---SKIPATTTQVSTTTQ- 916
Db 2817 -DDXTTKIPCKSSPELEDATSSRRRPRTRAQKVEVEEELAVLAVGLKLTQTSGETTHDKEP 2875
QY 917 -----DTTPPKITLITLILAPKVTITTKITITTEIANKPEETAKPKDRA----- 961
Db 2876 VEGSKGTAKF-----QPAKRNVDADVDIGSRQRPAPKEKAQPLEDLASFG 2922
QY 962 -----TNSKATTPKPKPTK-----APK-KPT----- 982
Db 2923 ELSQTPGHTHEELANGAADSFTSAPKQTPDGKPLKISRRVLRAPKVPFVGDVSTRDPVK 2982
QY 983 -----STKKPKTMRVSKPKT----- 998
Db 2983 SQSKSNTSLPLPKRGGGKDGSVTGKRLRCWPAPBEIVEELPASKKQVAPRAKGS 3042
QY 999 TPT---PRKMTSTPELNPTSRIAEAMLQTTTRPNQ-----TPNSKLVENVKPSDEAGGAE 1051
Db 3043 EPVIMKSLRTSAKRIEPAEELNSDMKTNKEHKLQDSVPENKGLSLRSRQDKTEAE 3102
QY 1052 GETHMLL-----RPHVFMVEVTPDMDYLPFRVFNQGIINP 1087
Db 3103 QQITEVFVLAERIEINRNEKKP-----MKTSPEND-----IQNP 3136

```

```

RESULT 15
A41819
proline-rich peptides 637K precursor, prostatic - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 13-May-1992 #sequence revision 13-May-1992 #text_change 09-Jul-2004
C:Accession: A42663; A41819; A31966; B20593; A20593
R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.
J. Biol. Chem. 267, 9884-9894, 1992
A:Title: The 4.4-kilodalton proline-rich polypeptides of the rat ventral prostate are the
le exon.
A:Reference number: A42663; MUID:92250652; PMID:1577819
A:Accession: A42663
A:Molecule type: DNA; mRNA
A:Residues: 1-3706; I, 3708-4077, F, 4079-4155, S, 4157-5761 <DE2>
A:Cross-references: UNIPROT:Q63455; GB:M86514
A:Experimental source: ventral prostate
A>Note: sequence inconsistent with the nucleotide translation
A>Note: sequence extracted from NCBI backbone (NCBI:100347, NCBI:100348)
R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.
submitted to the Protein Sequence Database, April 1992
A:Reference number: A41819
A:Accession: A41819
A>Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-5762 <DE1>
R:Hemschoote, K.; Peeters, B.; Dirckx, L.; Claessens, F.; De Clercq, N.; Heyns, W.; Wind
J. Biol. Chem. 263, 19159-19165, 1988
A:Title: A single 12.5-kilobase androgen-regulated mRNA encoding multiple proline-rich po
A:Reference number: A31966; MUID:89066721; PMID:3198617
A:Accession: A31966
A:Molecule type: mRNA
A:Residues: 3372-3540 <HEM>
A:Cross-references: GB:M20721; GB:J04188; NID:G206397; PIDN:AAA41950.1; PID:G554494
R:Peeters, B.; Heyns, W.; Bossyns, D.; Rombauts, W.
J. Biol. Chem. 258, 14206-14211, 1983
A:Title: Proline-rich polypeptides bound to rat prostatic binding protein. The primary st
A:Reference number: A94675; MUID:84061859; PMID:6685733
A:Accession: B20593
A:Molecule type: protein
A:Residues: 2020-2057 <PEE>
A>Note: this peptide, designated proline-rich polypeptide V, can be found at several loc
A:Accession: A20593
A:Molecule type: protein
A:Residues: 2822-2859 <PR2>
A>Note: this peptide, designated proline-rich polypeptide IV, can be found at several loc
C:Genetics:
A:Introns: status absent
A>Note: single copy gene with no introns
C:Superfamily: rat prostatic proline-rich peptides 637K precursor
C:Keywords: prostate; tandem repeat
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-5761/Product: prostatic proline-rich peptides 637K proprotein #status predicted <MAI
Query Match 14.3%; Score 833.4; DB 2; Length 5762;
Best Local Similarity 13.4%; Pred. No. 1e-08;
Matches 428; Conservative 144; Mismatches 419; Indels 2196; Gaps 115;
QY 4 KTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNDYNCQ----- 51
456 ETLPIHEI-----QPTQNEYGYQLPNVTVRPVDVA 486
Db 52 -----HYMECPDFKRVCTAELSCKRCFSEFGRZ----- 83
QY 487 LTVTSEPVKETESFLAPQFFVHALEYSND-----VEFFVNEEPPVQAP 531
Db 84 -----CCDDAQCKKYDK----- 95
QY 532 ETPGESQFESQLVEPQAQATEVDEBPKTSATSEQLAQFFPENDEVTVLPNHYQAQHSILS 591
QY 96 -----CCPYESFCAE----- 106
Db 592 NVTDQPLDLITITEKPMEMGTSPVYDAAAPAEVEFLSDQQGLVLSQSLPLEYLSLQ 651

```


QY 107 -----EHSYSEN----- 113
 Db 652 OBYTTGISQISEGGEFFTOETPEHSHVGMTEVEAQPQGHHEVTGPPLGHGVHSPALQ 711
 QY 114 -----QSSSSSSSSSSSTIW----- 130
 Db 712 NILTOYSTFPEKEQSPVGLGVPGHLEBFSVEPSPSQENSAMHVSVDMLFSLPDLQTF 771
 QY 131 ----KITSSKNSANRELQKLVKDNKNRKKKTPKPPV-----VDEA 172
 Db 772 RSTQSKSYKTIKHEDALTI-----TPEFSLDGSILFPQDLDLQIDST 817
 QY 173 GSG-----LDNGDFKVTTPDTS-----TTOHNKVTSP----- 200
 Db 818 GOGERSHIKSTLSKPPYVSNTKSAFOETMSETTVNSKQVDSLSTHLKTSSELPPDYTMG 877
 QY 201 ----KITTKAP-----RPSL-----PNSDTSKETS----- 225
 Db 878 LEPSLYQITQSSPKSLHENPKSFVLKPSAAQALKPRKEXLSPTNMIPHPSKPLKN 937
 QY 226 ----LTYNKETTIVETKETT-----TTNKQ 245
 Db 938 MYTHLPAHKVTPRQIOEQOGEYTTISSNGSFQPLDLEVTLSGIIPEVKHILPKETVNPQ 997
 QY 246 TSDGKEKTTSAKETQ--SIEKTSADLAPTSKVLAKTPK-----AETTK--GPAL 294
 Db 998 TYSQVKISHSQHVETQHPNSETTVQPLDLEFAINLQPTPKENFAQTQDITTMIGP-- 1055
 QY 295 TTPKE-----PPTTPK 306
 Db 1056 --PKEVIAQPEHHEGTIPIQDQAEYSTLTVSFQPLDQELTITSAREPFPHTVPQ 1113
 QY 307 EPASTTPKE-----TTPKE-----TIKSAPTPKEPAPTTTKSAPT 341
 Db 1114 QIIIVHPPEHPLVHSEQVYVQHPNTEALIQPLDLELTIPTQTAEGE-LPQTLQDST 1172
 QY 342 TPKEPAPTTKEPAP-----TTPKE-----PAPT-----TTKE----- 369
 Db 1173 QIIIEP-PTVVVGVPVIEEVTQTTSDQAEYPPSPTVSFQSLDLELITSPEATRESVHP 1231
 QY 370 -----PAPTTKSAPTPK- 383
 Db 1232 SLLQQTIVNPPHPLVHSEQVHTQHPNLTEATVQPLDPLTIIPQPTTEGELPQTLQD 1291
 QY 384 -----EPAPTTPKPAPTP-----KEP- 401
 Db 1292 STSQIIIEP-----PIKWVALPVYQVSVQDQAEYTTSTVSFQPLDQELTI-SEAIREFH 1347
 QY 402 APTTPKE-----PPTP-----TPKEPA-----PTTPKEP 424
 Db 1348 HPTVPOQTIVHPTKPLVHSEQVHTQHPNTEATVQPLDLELTIPTQTAEGELPQTLQD 1407
 QY 425 APTTPKEPAPTKAPTPKEPAPTPKE-----PAPT-----TKEPSPT- 467
 Db 1408 STQIIIEP-PTVVVGVPVIEEVTQTTSDQAEYPPSPTVSFQSLDLELTIPTPEPTEHF 1466
 QY 468 -----TPKE-----PAPT-----TTKSAP 481
 Db 1467 ITOKTIVFPMPYDVTLPQVSVQHLKPTREGIVQPLDLELTIPTQPTPEGELSQTVOEST 1526
 QY 482 TTKKEP-----APTTPKSAPTP-----KEPSP-----TTTKEPA----- 511
 Db 1527 TQNKEPKEVAVPVVYCAVTVPVTPSQYQAEYQKSLQPLDLELTVISEPKEAYHSTISK 1586
 QY 512 -----PTT-----PKE 517
 Db 1587 NSLAINQYVHIQHPNPAEATVQPLDLELTISSLOPTABGELLYSMQETVTOISEPPKQ 1646
 QY 518 -----PAPTTPKKAP-----TTPKEPAPT 537
 Db 1647 VVTPVPEQEVAVPAPVQDQAKYPLSSIVLSNLSLDQELTSSLELLGEAHLTIPTDETMVL 1706

QY 538 TPK-----EPAP 544
 Db 1707 PPXDRQGIYVDHDKHKLNLTEVNTQPFHLEHTVQHQPTIEERSQSIQKKTITQITEPG- 1765
 QY 545 TTTTKAPAPKAPAPTP-----KETAPTP-----KKLT----- 575
 Db 1766 ----KKVVPLOAESSEVTIPWILKETAFTPTPHSMALQSLDEKLTIIHSHSPGWTQOAHANLK 1822
 QY 576 ----- 575
 Db 1823 ESKGHTTGKILLDYABPNMBIELKHGHLFLTKTEATTESENTNOMTKSLKQVLTALFTON 1882
 QY 576 -----PTTPEKL----- 582
 Db 1883 KKSMLPALVESQDESOPPPNMSLOPLDQELTSSQPHGVPHIPNTEPEKIYLHYABPPTG 1942
 QY 583 ----APTTPKAPPTT-----PEELAPTTPE-----PPTTPEEPAP-- 616
 Db 1943 PFVEPPPELFLKTTKSRPVOGTATQMAASPKEVSRAPENKEAVLSGPGDQDESPPN 2002
 QY 617 -----TTPKAAAPNTPK-----EPAPTP----- 635
 Db 2003 MSLQSLDQELTSSQPHGVPHIPNTHGKIYLHYABP-PTGPFVEPPDLFFLTKTTSKPV 2061
 QY 636 ----KEPAPTPK----- 644
 Db 2062 EWTLTRDKSRKEMVSSQPKVEEAVLPHVHGQGESRSPNMSLSQLEQELTSSQPHGW 2121
 QY 645 ----EPAPTP-----KETAPTPKGTAPTTLKEP----- 670
 Db 2122 VPHPNTHGKIYLHYABP-PTGPFVEPPDLFFLRTTKSKPVQGTATRMVSPPEMWSLDP 2180
 QY 671 ----APTTPKPP----- 678
 Db 2181 ENKEAVFPAQGEKGESPPSNMSLSQSLDHELFMSSQPHGWIHPPTKPKIYLHYABEP 2240
 QY 679 ----APKELAPTTK----- 689
 Db 2241 TGPVPEPPDLFFLRTTKSKPVQGTTEMAXSPKEMVSGTPEYKEAVLSGPGDQDESPP 2300
 QY 690 ----EPT-----STTSDK 698
 Db 2301 PNTSLKSLDQEVAMSSQPHGVPHPPKTPGKIYLHSTIEPPPPPVKPTDLILVKTITKSK 2360
 QY 699 PAPTPK-----GTAPTTPKPA----- 716
 Db 2361 PAEWTPRDIRDKLKMVPHSPYEBAVFPAGHGEQDESPPNMPLOPLDQELTSSQPH 2420
 QY 717 ----PPTP-----KEPAPTPKGTAP-----TTLKEPAPTTPKKPA--PKELA-- 753
 Db 2421 GWVPHHPNTPGKIYLHYABP-PTGPFVEPPDLFFLRTTKSKPVQGSPEIAKSPKEMVSQ 2479
 QY 754 ----PTTTKG----- 759
 Db 2480 TPEYKEAVLSGPGDQDESPPSNMSLSQSLDQEVMTSSQPHSGVPHPPKTPGKIYLHSTIE 2539
 QY 760 ----PT-----STTSDKAPAPTP-----KETAPTTPE-----PA----- 785
 Db 2540 PPGPFVXPTDLILVKTITKSPAEWTPRRIDKLLKEMVPHSPYEBAVFPAGHGEQDES 2599
 QY 786 ----PPTPKP-----APTTPETP-----PPTTSEVSTP-----TTTK 814
 Db 2600 GSPPNNPLOPLDQELTSSQPHGVPHHPNTPGKIYLHYABPPTGPFVEPPDLFFLRTPK 2659
 QY 815 E-----PTTIHKSPE-----STPELS----- 831
 Db 2660 SKFVGTGTQMAKSPPEMVSLSPKNKTVFPAGKQDESPPNMSLSQSLDQELTSSQ 2719
 QY 832 ----AEP-----TPKALENSPK-- 845
 Db 2720 PHGWPHPNTHGKIYLHYABPPTGPFVEPPDLFFLTKTKSKPMQGSQPRQDKSPREMT 2779
 QY 846 ----FGVP----- 849

```
Db 2780 QSPYEESLLPAHAEGQESRAPHESLQPLDQELSLSSHPHGWIPHHNTPDKIYLHYA 2839
Qy 850 -----TTKTPAATKP-EMTTAKDKTTERDLRTTPTTTAA----- 884
Db 2840 EPPTGPFVEPPDLFFLTKTK-----SKVQGTATKTDKSPEDRVSTQPYEKEAVLSGPGED 2895
Qy 885 -----PKMT-----KETATITE----- 896
Db 2896 QDESPPPNMSLQSLDOELAISSQPHGWIPHSFNAPDKIYLHYABPTGPFVEPPDLFFL 2955
Qy 897 KITESK-ITATTQTSTTTQ----- 916
Db 2956 KITKSKPLQGTPTQMAKSPKEMVSTQPYEKEADLSAPGENRDESSPPNMSLHPLDQELS 3015
Qy 917 -----DTTPF-----KITTLKTTLAPKVTTTKTITTTE 946
Db 3016 LSSQPHGWIPHPNTHGKIYLHYABPTGPFVEPPDLFFLTKTKSP----- 3062
Qy 947 INMKPEETAK-PKD----- 959
Db 3063 VQGSFQIDKSPKEVFSQSPSESVLPAQAEQESRAPPHMSLQLLDQKLTLSHPHG 3122
Qy 960 -----RATNSKATTPKPKTKAPK----- 980
Db 3123 WIPHPNTPGKIYLHYABPTGPFVEPPDLFFLTKTKSPQVQSPQVQDKSPKEMFTQSP 3182
Qy 981 -----PTSTKK-----PK 988
Db 3183 EYESVLPAAEQEGESRAPPHMSLQPLDQDLTSSHPHGWIHPNTPDKIYLHYAEP 3242
Qy 989 TMRVRKP-----KTTP-----PRKMTSTMP-----LNP 1014
Db 3243 TGPFFVEPPDLFFLRTTKGPAQWTPQMAKSPPEMVSLSPENKENVFPQAGQEGESISP 3302
Qy 1015 TSRIAEAMLQTT-----RPN-----QTPNSKL 1037
Db 3303 PHMSLQPLDQDLTPSSHPHGWIHPNTHGKIYLHYABPTGPFVEPPDLFFLTKTKSP 3362
Qy 1038 VEVNP----- 1042
Db 3363 VQVSPRQIDKSPKEVFTQSPVEESVLPAAEQEGESRAPPHMSLQPLDQDLTSSHPHG 3422
Qy 1043 -----KSEDAG----- 1049
Db 3423 WITHHPNTHGKIYLHYABPTGPFVEPPDLFFLTKTKSPVHGSFRQIDKSHKEMFTQSP 3482
Qy 1050 -----AEGE-----TPHMLLR-----PHVEMPE--VTPDMDYLPRV-PN 1080
Db 3483 EYESVLPAAEQEGESRAPPHMSLQPLDQDLTSSHPHGWIHPNTPDKIYLHYAEP 3542
Qy 1081 QGIINP 1087
Db 3543 TGPFFVEP 3549
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 88.5238 Seconds
(without alignments)
4449.477 Million cell updates/sec

Title: SEQ1-C
Perfect score: 5887
Sequence: 1 MAWKTLPIYLLLSVFVQI.....DMDYLPVPNGQIIINPMLS 1098

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:.*
1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5872.8	99.8	✓ 1299	4 AAM24322	Human EST
2	5872.8	99.8	1404	2 AAB26049	MSP precu
3	5872.8	99.8	1404	4 AAB29773	Human meg
4	5872.8	99.8	1404	4 AAB60568	Human meg
5	5872.8	99.8	1404	8 ADM98014	Human meg
6	5872.8	99.8	1404	4 AAB32262	Novel hum
7	5863.8	99.6	1404	7 ADK65839	Angiogene
8	5607.9	95.3	1311	8 ADK67912	Human ext
9	5598.8	95.1	1320	7 ADK65819	Angiogene
10	5343.8	90.8	1270	8 ADK67911	Human ext
11	2920	49.6	546	4 ABUS3252	Human tes
12	2888	49.1	551	4 ABUS3253	Human tes
13	2880.1	48.9	902	4 AAB29778	Human MSF
14	2757	46.8	513	4 ABUS3254	Human tes
15	2222.8	37.8	452	4 AAR80041	Human meg
16	1866.8	26.6	292	5 AAU11261	Human HAP
17	1409.9	23.9	538	5 AAOL18834	3' cartil
18	1396.9	23.7	5179	4 AAM24516	C899P pre
19	1396.9	23.7	5179	6 ABP55365	Human col
20	1396.9	23.7	5179	6 ABO07258	Human p53
21	1396.9	23.7	5179	7 AD448091	Human Pro
22	1396.9	23.7	5179	7 AD444998	Human Pro
23	1132	19.2	1664	2 AAW43106	C. thermo
24	1127.3	19.1	8991	4 ABO08487	S. pneumo
25	1100.7	18.7	1795	4 ABB69806	Drosophil

26	1100.4	18.7	717	4 ABUS3144	Human tes
27	1097	18.6	214	4 ABUS3255	Human tes
28	1067.7	18.1	763	3 AAG38942	Arabidops
29	1048.4	17.8	1049	4 AB611364	Drosophil
30	1047.2	17.8	2284	4 AB371434	Drosophil
31	1032.6	17.5	5703	8 AD23265	Human MUC
32	1009.9	17.2	188	5 AAOL18833	5' cartil
33	995.3	16.9	4315	5 ABP43908	MUC5B par
34	987.9	16.8	770	4 ABUS3141	Human tes
35	981	16.7	778	4 ABUS3143	Human tes
36	975	16.6	2768	4 AB68397	Drosophil
37	974.4	16.6	1892	4 ABUS3155	Human tes
38	968.2	16.4	1480	7 ADN39110	Cancer/an
39	967.7	16.4	1538	7 ADI21202	Novel hum
40	965.9	16.4	2112	4 ABB60403	Drosophil
41	964.1	16.4	717	4 ABUS3150	Human tes
42	964.1	16.4	717	4 ABUS3149	Human tes
43	964.1	16.4	717	4 ABUS3151	Human tes
44	964.1	16.4	717	4 ABUS3145	Human tes
45	964.1	16.4	717	4 ABUS3148	Human tes

ALIGNMENTS

RESULT 1
AAM24322
ID AAM24322 standard; protein; 1299 AA.
XX
AC AAM24322;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST encoded protein SEQ ID NO: 1847.
XX
KW Human, sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;
KW gene therapy; nutrition.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US002687.
XX
PR 25-JAN-2000; 2000US-00491404.
PR 17-JUL-2000; 2000US-00617746.
PR 03-AUG-2000; 2000US-00631451.
PR 15-SEP-2000; 2000US-00663870.
XX
PA (HYSE-) HYSEQ INC.
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
DR WPI: 2001-476164/51.
DR N-PSDB: AAH98981.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use.
XX
PS Claim 20; Page 1198-1201; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics. Gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a

XX	Claim 1, 2 and 3; Fig 1; 87pp; English.	
PS	The sequence given is a full length translation from the megakaryocyte	
XX	stimulating factor (MSF) precursor. The sequence covered by exons II, III	
CC	and IV encodes megakaryocyte stimulating factor (MSF). This sequence is	
CC	modified by the addition of an N-terminal sequence encoding a secretory	
CC	leader, an initiating methionine preceding exon II and a terminating	
CC	codon following exon IV. The cDNA sequence given contains sequences	
CC	derived from human megakaryocyte colony stimulating factor (meg-CSF).	
CC	Exon I contains the initiating methionine, and encodes a classical	
CC	mammalian protein secretion signal sequence. The sequence encoding the	
CC	original meg-CSF includes exons II-IV and is thought to terminate in the	
CC	region between amino acid residues 134 - 147. The primary transcript of	
CC	this gene may be cleaved in different ways to yield a family of mRNAs	
CC	each encoding a different MSF protein. Exons V and VI are thought to be	
CC	related to the activity of the factor and are also implicated in the	
CC	stability, folding and processing of the molecule. These exons are also	
CC	thought to play a role in the observed synergy of MSF with other	
CC	cytokines. Exons V - XII are believed to be implicated in the processing	
CC	or folding of the appropriate structure of the resulting factor, i.e. one	
CC	or more of these exons may contain sequences which direct proteolytic	
CC	cleavage, adhesion, organisation of the cellular matrix or extracellular	
CC	matrix processing. Both naturally occurring and non-naturally occurring	
CC	MSF's may be characterised by various combinations of alternatively	
CC	spliced exons from this sequence, with the exons spliced together in	
CC	different orders to form different members of the MSF family. (Updated on	
CC	25-MAR-2003 to correct PN field.)	
XX		
SQ	Sequence 1404 AA;	
	Query Match 99.8%; Score 5872.8; DB 2; Length 1404;	
	Best Local Similarity 96.3%; Pred. No. 3.2e-154;	
	Matches 1098; Conservative 0; Mismatches 0; Indels 42; Gaps 1;	
QY	1 MAWKTLPIYLLLLSVFIQVSSODLSCAGRCGEGYSDRATCNCYNCOHYMECCPDF 60	
DB	1 MAWKTLPIYLLLLSVFIQVSSODLSCAGRCGEGYSDRATCNCYNCOHYMECCPDF 60	
QY	61 KRVTAEALSCGRCPESPERGECDCDAQCKYDCCPDYSFCAEVHNPTSPSSKKAP 120	
DB	61 KRVTAEALSCGRCPESPERGECDCDAQCKYDCCPDYSFCAEVHNPTSPSSKKAP 120	
QY	121 PPSGASQTIKSTKSPKPNKKTKVIESEITE----- 156	
DB	121 PPSGASQTIKSTKSPKPNKKTKVIESEITE----- 180	
QY	157 -----KVQDKNKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 198	
DB	181 KTSSQNSAANRELQCKLVKDNKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 240	
QY	199 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTETTKTSTG 258	
DB	241 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTETTKTSTG 300	
QY	259 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKEPAS 318	
DB	301 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKEPAS 360	
QY	319 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 378	
DB	361 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 420	
QY	379 APTTTKSAPTTKPEAPTTPKKAPTTKPEAPTTKPEPTTTPKEAPTTKPEAPTTTK 438	
DB	421 APTTTKSAPTTKPEAPTTPKKAPTTKPEAPTTKPEPTTTPKEAPTTKPEAPTTTK 480	
QY	439 EPAPTAKPKPAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPT 498	
DB	481 EPAPTAKPKPAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPT 540	
QY	499 TTKSAPTTKPEPSPTTTTKPEAPTTPKKAPTTKPEAPTTKPEAPTTTKPEAPTTTKKP 558	

541 TTKSAPTTKPEPSPTTTTKPEAPTTPKKAPTTKPEAPTTKPEAPTTTKPEAPTTTKKP 600

559 APTAPKPEAPTTKPEAPTTKPKLTPTTPEKLAAPTTPEKAPTTPELAAPTTPEEBPTPT 618

601 APTAPKPEAPTTKPEAPTTKPKLTPTTPEKLAAPTTPEKAPTTPELAAPTTPEEBPTPT 660

619 PEEPAAPTTKAAAPNTPEKAPTTKPEAPTTKPEAPTTKPEAPTTPKETAPTTKGTAPTTLKEP 678

661 PEEPAAPTTKAAAPNTPEKAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTPKETAPTTLKEP 720

679 APTTPKKAPKELAPTTKEPTSTTSKAPTTKGTAPTTKGTAPTTKGTAPTTKGTAPTTKGT 738

721 APTTPKKAPKELAPTTKEPTSTTSKAPTTKGTAPTTKGTAPTTKGTAPTTKGTAPTTKGT 780

739 TAPTTLKEAPTTKPKAPKELAPTTKGTSTTSKAPTTKGTAPTTKGTAPTTKGTAPTTKGT 798

781 TAPTTLKEAPTTKPKAPKELAPTTKGTSTTSKAPTTKGTAPTTKGTAPTTKGTAPTTKGT 840

799 KPAPTTPETPPPTTSEVSTPTTKEPTTHKSDSTPELSAETPKALENSKEPGVPT 858

841 KPAPTTPETPPPTTSEVSTPTTKEPTTHKSDSTPELSAETPKALENSKEPGVPT 900

859 TKTPAATKPEMTTAKDKTTERDLRTPTTAAAPKTKETATTTTEKTTESKITATTQV 918

901 TKTPAATKPEMTTAKDKTTERDLRTPTTAAAPKTKETATTTTEKTTESKITATTQV 960

919 TSTTTQDTPFKITTLKTTTLAPKVTITTKITTTIMNKPEBTAKPKDRATNSKATTPK 978

961 TSTTTQDTPFKITTLKTTTLAPKVTITTKITTTIMNKPEBTAKPKDRATNSKATTPK 1020

979 PQKPTKAPKKTSTKXPKTMPRVKPKTPTPRKMTSTMPKMTSTMPKMTSTMPKMTSTMPKMT 1038

1021 PQKPTKAPKKTSTKXPKTMPRVKPKTPTPRKMTSTMPKMTSTMPKMTSTMPKMTSTMPKMT 1080

1039 QTNSKLIVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMDYLPRVNPQGIINPMLS 1098

1081 QTNSKLIVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMDYLPRVNPQGIINPMLS 1140

RESULT 3

AAB29773

ID AAB29773 standard; protein; 1404 AA.

XX AAB29773;

XX AC AAB29773;

DT 28-FEB-2001 (first entry)

XX Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.

DE Human MSF, megakaryocyte stimulating factor; tribonectin;

KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;

KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;

KW friction coefficient reduction; gene therapy; antiarthritic; osteopathic.

XX Homo sapiens.

OS WO200064930-A2.

XX PN 02-NOV-2000.

XX PD 24-APR-2000; 2000WO-US010953.

XX PF 23-APR-1999; 99US-00298970.

XX PR (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

PA Jay GD;

XX WPI; 2001-024673/03.

XX DR N-PSDB; AAC81498.

XX Novel tribonectin polypeptide useful as lubricant for treating

PT osteoarthritis, comprises O-linked lubricating moiety.

PT

XX PS Claim 3; Page 7; 47pp; English.

CC The invention relates to a human tribonectin which is a product of

CC alternative splicing of the human MSF (megakaryocyte stimulating factor)

CC gene. The tribonectin has at least one O-linked oligosaccharide

CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats

CC of a motif having at least 50% identity to the sequence KEPAPTT

CC (AAB25774). The invention also relates to a nucleic acid encoding a human

CC MSF-derived tribonectin; a biocompatible composition comprising a human

CC tribonectin for inhibiting tissue adhesion formation; and a method of

CC diagnosing osteoarthritis or a predisposition to osteoarthritis by

CC measuring the amount of MSF or its fragment in a biological sample of a

CC mammal, wherein an increased amount of MSF compared to a control

CC indicates the presence of or predisposition to developing osteoarthritis.

CC The tribonectin and DNA encoding it are useful in the treatment of

CC osteoarthritis, where they may be used for lubricating mammalian joints,

CC such as articulating joints of humans, dogs or horses. The tribonectin,

CC when formulated as a membrane, foam, gel or fibre, is useful for

CC inhibiting adhesion between two surfaces such as the injured tissues of a

CC mammal, where the injury is caused by a surgical insertion or trauma, or

CC an artificial device e.g., an orthopaedic implant. In particular, one of

CC the surfaces is pericardial tissue. DNA encoding a tribonectin may be

CC used in gene therapy. The present sequence represents human MSF

XX Sequence 1404 AA;

Query Match 99.8%; Score 5872.8; DB 4; Length 1404;

Best Local Similarity 96.3%; Pred. No. 3.2e-154;

Matches 1098; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSODLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60

DB 1 MAWKTLPIYLLLLSVFVIQVSSODLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60

QY 61 KVVCTAELSCGRCPESFERGECDDAQQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120

DB 61 KVVCTAELSCGRCPESFERGECDDAQQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120

QY 121 PPSGASQIKSTTKSPKPPNKKTKVIESEITE----- 156

DB 121 PPSGASQIKSTTKSPKPPNKKTKVIESEITEHSVSENQESSSSSSSSSSSTIW 180

QY 157 -----KVQDNKNRTKKPKPPVVDVDEAGSLDNGDFKVTPTDTST 198

DB 181 KIKSSKNSAANRELQKKLVKDNKNRTKKPKPPVVDVDEAGSLDNGDFKVTPTDTST 240

QY 199 TQHNKVSTSPKITTAKPINRPSLPNNSDTSKETSLVNKEVETKETTNNKQSTDG 258

DB 241 TQHNKVSTSPKITTAKPINRPSLPNNSDTSKETSLVNKEVETKETTNNKQSTDG 300

QY 259 KEKTSKETSQIKSTSAKDLAPTSKVLAKPTPKAETTTKGPALTTTKEPTTTPKEPAS 318

DB 301 KEKTSKETSQIKSTSAKDLAPTSKVLAKPTPKAETTTKGPALTTTKEPTTTPKEPAS 360

QY 319 TTPKEPTTTIKSAPTTTKEPAPTTKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 378

DB 361 TTPKEPTTTIKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 420

QY 379 APTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 438

DB 421 APTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 480

QY 439 EPAPTAPKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 498

DB 481 EPAPTAPKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 540

QY 499 TTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 558

DB 541 TTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 600

QY 559 APTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 618

DB 601 APTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 660

QY 619 PEEPAPTTTKAAAPNTPKAPPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPT 678

DB 661 PEEPAPTTTKAAAPNTPKAPPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPT 720

QY 679 APTTPKAPKAPKELAPTTTKEPSTTSDKAPPTTPKGTAPTTTPKEPAPTTTPKEPAPTTPKG 738

DB 721 APTTPKAPKAPKELAPTTTKEPSTTSDKAPPTTPKGTAPTTTPKEPAPTTTPKEPAPTTPKG 780

QY 739 TAPTTTKEPAPTTTPKAPKAPKELAPTTTKEPSTTSDKAPPTTPKGTAPTTTPKEPAPTTTPK 798

DB 781 TAPTTTKEPAPTTTPKAPKAPKELAPTTTKEPSTTSDKAPPTTPKGTAPTTTPKEPAPTTTPK 840

QY 799 KPAPTTTPETPTTSEVSTPTTTPKEPTTIHKSPDSESTPELSAETPKALENSPKPGVPT 858

DB 841 KPAPTTTPETPTTSEVSTPTTTPKEPTTIHKSPDSESTPELSAETPKALENSPKPGVPT 900

QY 859 TKTPAATKEMTTAKDKTTERDLRTTPETTTAAPQWTKETATTTKTESKITATTTQV 918

DB 901 TKTPAATKEMTTAKDKTTERDLRTTPETTTAAPQWTKETATTTKTESKITATTTQV 960

QY 919 TSTTTQDTPPFKITTLLKTTTLAPKVTITTKTITTTIMNKPBETAKPKDRATNSKATTPK 978

DB 961 TSTTTQDTPPFKITTLLKTTTLAPKVTITTKTITTTIMNKPBETAKPKDRATNSKATTPK 1020

QY 979 POKPTKAPKPTSTKPKTMPRVKPKTPTPRKMTSTWPELNPTSRIAEAMLOTTTSPN 1038

DB 1021 POKPTKAPKPTSTKPKTMPRVKPKTPTPRKMTSTWPELNPTSRIAEAMLOTTTSPN 1080

QY 1039 QTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFPVETPDMDYLPVFNQGIINPMLS 1098

DB 1081 QTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFPVETPDMDYLPVFNQGIINPMLS 1140

RESULT 4

ID AAB60568 standard; protein; 1404 AA.

XX AC AAB60568;

XX DT 27-APR-2001 (first entry)

XX DE Human megakaryocyte stimulating factor (MSF, CACP).

XX KW Human; CACP protein; camptodactyly-arthropathy-coxa vara-pericarditis;

XX KW MSF; megakaryocyte stimulating factor; synovial lubricant;

XX KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;

XX KW antiarthritic.

OS Homo sapiens.

PN WO200107068-A1.

XX PD 01-FEB-2001.

XX PF 21-JUL-2000; 2000WO-US0200002.

XX PR 23-JUL-1999; 99US-0145328P.

XX PR 19-JUL-2000; 2000US-00145328.

XX PA (UYCA-) UNIV CASE WESTERN RESERVE.

XX PI Warman ML;

XX DR WPI; 2001-182721/18.

XX PT New composition comprising the camptodactyly-arthropathy-coxa vara-

XX PT pericarditis protein in combination with an anesthetic, useful for

XX PT treating osteoarthritis, or as lubricants of tissue and joints.

XX PS Example 1; Page: 34pp; English.

The invention relates to a method of treating osteoarthritis via the administration of a composition comprising the camptodactyl-arthropathy-coxa vara-pericarditis (CACP) protein, or portions of the CACP protein. The composition may further comprise a local anaesthetic. The composition of the invention may be administered via intra-articular or intravenous injection. The human CACP protein is identified in the invention as being megakaryocyte stimulating factor (MSF). The gene encoding CACP protein (MSF) is located on chromosome 1q25-31, and mutations in this gene are responsible for the heritable disorder camptodactyl-arthropathy-coxa vara-pericarditis, in which patients have synovial hyperplasia without evidence of inflammation. CACP protein (MSF) acts as a synovium lubricant, and can be used to lubricate tissue and joints in the treatment of osteoarthritis. The composition may be applied to reduce the symptoms of osteoarthritis (e.g., joint pain, loss of range of movement or joint damage). The present sequence represents human megakaryocyte stimulating factor (MSF, CACP protein). Note: This sequence is not given in its entirety in figure 4 of the specification, although a GenBank accession number was given. This sequence was therefore obtained from GenBank (U70316)

XX
XX
SQ

Sequence 1404 AA;

Query Match 99.8%; Score 5872.8; DB 4; Length 1404;
Best Local Similarity 96.3%; Pred. No. 3.2e-154;
Matches 1098; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MAWKTLPIYLILLVSVFIQVSSODLSACGACGCGYSRATCNCDYNCOHMECCPDF 60
DB 1 MAWKTLPIYLILLVSVFIQVSSODLSACGACGCGYSRATCNCDYNCOHMECCPDF 60

QY 61 KEVCTAELSCGRCFESFERGECDCDAQCKYKCCPDYSEFCAEVHNPTSPSSKKAP 120
DB 61 KEVCTAELSCGRCFESFERGECDCDAQCKYKCCPDYSEFCAEVHNPTSPSSKKAP 120

QY 121 PPSGASQTIKSTKRSPPNKKTKVIESBEITE----- 156
DB 121 PPSGASQTIKSTKRSPPNKKTKVIESBEITE----- 156

QY 157 -----KVKNKKNRTKKKTPPPVVDAGSLDNGDFKVTTPDTST 198
DB 181 KTKSSKNSANRELQKLVKDNKKNRYKKETPPVVDAGSLDNGDFKVTTPDTST 240

QY 199 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKTETVTKTTNKQTSYDG 258
DB 241 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKTETVTKTTNKQTSYDG 300

QY 259 KEKTSKETSQSIKTSADLAPTSKVLAKPTKAEITTKGPALTPKPEPTTPKEPAS 318
DB 301 KEKTSKETSQSIKTSADLAPTSKVLAKPTKAEITTKGPALTPKPEPTTPKEPAS 360

QY 319 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 378
DB 361 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 420

QY 379 APTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEPTTPKPEAPTTKPEAPTTKPE 438
DB 421 APTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEPTTPKPEAPTTKPEAPTTKPE 480

QY 439 EPAPTAPKKPAPTTKPEAPTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPTT 498
DB 481 EPAPTAPKKPAPTTKPEAPTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPTT 540

QY 499 TTKSAPTTTKPEAPTTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKKP 558
DB 541 TTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTKPEAPTTTKKP 600

QY 559 APTAPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTPEELAPTTPEEPTPTT 618
DB 601 APTAPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTPEELAPTTPEEPTPTT 660

QY 619 PPEAPTTTKAAAPNTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 678
DB 661 PPEAPTTTKAAAPNTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 720

QY 679 APTPKKAPKELAPTTTKPEPTSTTSKDPAPTTPKGTATTTKPEAPTTKPEAPTTKPG 738
DB 721 APTPKKAPKELAPTTTKPEPTSTTSKDPAPTTPKGTATTTKPEAPTTKPEAPTTKPG 780

QY 739 TAPTTKPEAPTTKPKAPKELAPTTTKGTSTTSKDPAPTTPKGTATTTKPEAPTTKPK 798
DB 781 TAPTTKPEAPTTKPKAPKELAPTTTKGTSTTSKDPAPTTPKGTATTTKPEAPTTKPK 840

QY 799 KPAPTTPEPTTSPVSTPTTKETTHKSPDESTPDSABPTPKALENSKPEGVPT 858
DB 841 KPAPTTPEPTTSPVSTPTTKETTHKSPDESTPDSABPTPKALENSKPEGVPT 900

QY 859 TKTPAAKPEMTTAKDKTERDLRTTPTTAAAPKMTKETATTTTEKTTESKITATTTOV 918
DB 901 TKTPAAKPEMTTAKDKTERDLRTTPTTAAAPKMTKETATTTTEKTTESKITATTTOV 960

QY 919 TSTTQDTTTPFKITTLTKTTTLAPKVTTKKTTTTEIMNKPEETAKPKORATNSKATTPK 978
DB 961 TSTTQDTTTPFKITTLTKTTTLAPKVTTKKTTTTEIMNKPEETAKPKORATNSKATTPK 1020

QY 979 PQKPTKAPKPTSTKPKTMPRVKPKTTPTPEKMTSTMPELNPTTSRIEAMLOTTTRPN 1038
DB 1021 PQKPTKAPKPTSTKPKTMPRVKPKTTPTPEKMTSTMPELNPTTSRIEAMLOTTTRPN 1080

QY 1039 QTNSKLIVEVNPKSSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1098
DB 1081 QTNSKLIVEVNPKSSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1140

RESULT 5
ADM98014
ID ADM98014 standard; protein; 1404 AA.
XX ADM98014;
AC ADM98014;
XX
DT 01-JUL-2004 (first entry)
XX Human megakaryocyte stimulating factor (MSF).
DE lubricating polypeptide; O-linked oligosaccharide; joint lubrication;
KW CAP; camptodactyl-arthropathy pericarditis; osteoarthritis; human;
KW megakaryocyte stimulating factor; MSF.
XX Homo sapiens.
XX OS
XX US2004072741-A1.
XX 15-APR-2004.
XX 02-JUL-2001; 2001US-008997188.
XX 23-APR-1999; 99US-00298970.
XX 24-APR-2000; 2000US-00556246.
XX (JAYG/) JAY G D.
XX Jay GD;
PI WPI; 2004-373948/35.
XX DR N-PSDB; ADM98015.
XX
PT New tribonectin polypeptides and polynucleotides for lubricating joints
PT or other tissues to prevent or treat Camptodactyl-arthropathy-
PT pericarditis syndrome or osteoarthritis.
XX
PS Claim 1; SEQ ID NO 1; 34pp; English.
XX
CC The invention relates to a lubricating polypeptide and at least one O-
CC linked oligosaccharide. The composition and methods are useful for
CC lubricating joints or other tissues to prevent or treat camptodactyl-
CC arthropathy pericarditis (CAP) or osteoarthritis in mammals. The present
CC sequence represents the amino acid sequence of the human megakaryocyte

Db 12 MAWTLPIYLLLSVFIQVSSQDLSGAGCGEYSRDATCNCYNQHYMECCPDF 71
Qy 61 KVCCTAELSKGRCFESFERGECDDAQCCKYKCCPDYEFCAEVHNPTSPSSSKAP 120
Db 72 KVCCTAELSKGRCFESFERGECDDAQCCKYKCCPDYEFCAEVHNPTSPSSSKAP 131
Qy 121 PPASGASQIKSTTKSPKPPNKKTKKVISEBITE----- 156
Db 132 PPASGASQIKSTTKSPKPPNKKTKKVISEBITEHSHSVENQESSSSSSSSSTIW 191
Qy 157 -----KVQDNKNRUKKPTPKPPVVDGAGSLDNGDFKVTPTDST 198
Db 192 KIKSKNSAANRELQKLVKDNKNRUKKPTPKPPVVDGAGSLDNGDFKVTPTDST 251
Qy 199 TQHNKVSTSPKITTAKINPRSLPNSTSKETSLTVNKKETTVETKETTNNKQTSIDG 258
Db 252 TQHNKVSTSPKITTAKINPRSLPNSTSKETSLTVNKKETTVETKETTNNKQTSIDG 311
Qy 259 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTKGPAETTKPEPTTPKBPAS 318
Db 312 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTKGPAETTKPEPTTPKBPAS 371
Qy 319 TTPKEPTPTTIKSAETTPKEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEP 378
Db 372 TTPKEPTPTTIKSAETTPKEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEP 431
Qy 379 APTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEP 438
Db 432 APTTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEP 491
Qy 439 EPAPTAPKPPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTT 498
Db 492 EPAPTAPKPPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTT 551
Qy 499 TTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTT 558
Db 552 TTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTT 611
Qy 559 APTAKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEP 618
Db 612 APTAKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEP 671
Qy 619 PEPAPTTKAAAPNTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEP 678
Db 672 PEPAPTTKAAAPNTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEP 731
Qy 679 APTPKKAPKELAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEP 738
Db 732 APTPKKAPKELAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEP 791
Qy 739 TAPTTLKPEPAPTTKPKAPKELAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEP 798
Db 792 TAPTTLKPEPAPTTKPKAPKELAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEP 851
Qy 799 KPAPTTPEPTTSEVSTPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEP 858
Db 852 KPAPTTPEPTTSEVSTPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEP 911
Qy 859 TKTPAATKPEMTTAKDKTTERDLRTTPTTTAAAPKMTKETATTTKTESKITATTQV 918
Db 912 TKTPAATKPEMTTAKDKTTERDLRTTPTTTAAAPKMTKETATTTKTESKITATTQV 971
Qy 919 TSTTTQDTPFKITTLTKTTTLAPKVTTKTITTEIMNKPSETAKPKDRATNSKATTPK 978
Db 972 TSTTTQDTPFKITTLTKTTTLAPKVTTKTITTEIMNKPSETAKPKDRATNSKATTPK 1031
Qy 979 POKPTAKPKPTSTKPKTTPVRKPKTTPTPKMTSTMPKLNPTSRIAEAMLOTTTPN 1038
Db 1032 POKPTAKPKPTSTKPKTTPVRKPKTTPTPKMTSTMPKLNPTSRIAEAMLOTTTPN 1091
Qy 1039 QTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1098

Db 1092 QTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1151
RESULT 7
ADK65839
ID ADK65839 standard; protein; 1404 AA.
XX AC ADK65839;
XX DT 06-MAY-2004 (first entry)
XX DE Angiogenesis-differentially expressed protein #53.
XX KW cytostatic; cardiant; vasotropic; antiarteriosclerotic;
XX KW angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
XX KW gene expression; cancer; coronary artery disease; myocardial ischemia;
XX KW coronary arteriosclerosis; forensic medicine.
XX OS Homo sapiens.
XX WO2003066831-A2.
XX PD 14-AUG-2003.
XX PF 07-FEB-2003; 2003WO-US0003848.
XX PR 07-FEB-2002; 2002US-00067482.
XX PR 10-JUN-2002; 2002US-00164595.
XX PR 16-AUG-2002; 2002US-0403649P.
XX PR 03-JAN-2003; 2003US-0437746P.
XX (ORIG-) ORIGENE TECHNOLOGIES INC.
XX Sun Z, Li X, Kovacs KF, Fan W, Jay G;
XX WPI; 2003-731502/69.
XX DR
XX PT Determining the angiogenic index of a tissue or cell sample using
XX PT expression levels of differentially expressed genes, useful for
XX PT diagnosing or treating cancer, coronary artery disease, myocardial
XX PT ischemia and/or arteriosclerosis.
XX PS Disclosure; SEQ ID NO 78; 296pp; English.
XX CC
XX CC The invention relates to a method of determining the angiogenic index of
XX CC a tissue or cell sample comprising assessing, in a sample, the expression
XX CC levels of one or more differentially-expressed gene from any of 34 DNA
XX CC sequences, given in the specification, where the levels are indicative of
XX CC the angiogenic index. The methods and compositions of the present
XX CC invention are useful for diagnosing, preventing and/or treating cancer,
XX CC coronary artery disease, myocardial ischemia or coronary
XX CC arteriosclerosis. They can also be used in research, drug discovery and
XX CC forensic medicine involving angiogenesis. This sequence corresponds to
XX CC one of the differentially expressed proteins of the invention.
XX SQ
XX SQ Sequence 1404 AA;
Query Match 99.6%; Score 5863.8; DB 7; Length 1404;
Best Local Similarity 96.1%; Pred. No. 5.6e-154;
Matches 1096; Conservative 0; Mismatches 2; Indels 42; Gaps 1;
Qy 1 MAWTLPIYLLLSVFIQVSSQDLSGAGCGEYSRDATCNCYNQHYMECCPDF 60
Db 1 MAWTLPIYLLLSVFIQVSSQDLSGAGCGEYSRDATCNCYNQHYMECCPDF 60
Qy 61 KVCCTAELSKGRCFESFERGECDDAQCCKYKCCPDYEFCAEVHNPTSPSSSKAP 120
Db 61 KVCCTAELSKGRCFESFERGECDDAQCCKYKCCPDYEFCAEVHNPTSPSSSKAP 120
Qy 121 PPASGASQIKSTTKSPKPPNKKTKKVISEBITE----- 156
Db 121 PPASGASQIKSTTKSPKPPNKKTKKVISEBITEHSHSVENQESSSSSSSSSTIR 180

QY 157 -----KVNKKNRKTKKKPTKPPVVDVDEAGSLDNGDKFKVITPTDST 198
 Db 181 KIKSSKNSAANRELQKLVKVNKKNRKTKKKPTKPPVVDVDEAGSLDNGDKFKVITPTDST 240
 QY 199 TQHNKVSPTSPIITAKPINRPSLSPNSDTSKETSLSLVNKKETVETKTTTTNNKQTSDDG 258
 Db 241 TQHNKVSPTSPIITAKPINRPSLSPNSDTSKETSLSLVNKKETVETKTTTTNNKQTSDDG 300
 QY 259 KEKTTSAKETQSIKTSKADLAPTQKVLAKPTPKAETTTKGPALTTPKEPTPTPKBPAS 318
 Db 301 KEKTTSAKETQSIKTSKADLAPTQKVLAKPTPKAETTTKGPALTTPKEPTPTPKBPAS 360
 QY 319 TTPKEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 378
 Db 361 TTPKEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 420
 QY 379 APTTTKSAPTTKPEAPTTPKKAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAP 438
 Db 421 APTTTKSAPTTKPEAPTTPKKAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAP 480
 QY 439 EPAPTAPKAPPTTKPEAPTTKPEAPTTTKKESPTTKPEAPTTTKSAPTTTKPEAPT 498
 Db 481 EPAPTAPKAPPTTKPEAPTTKPEAPTTTKKESPTTKPEAPTTTKSAPTTTKPEAPT 540
 QY 499 TTKSAPTTKESPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKKP 558
 Db 541 TTKSAPTTKESPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKKP 600
 QY 559 APTAPKPEAPTTKPEAPTTTKKLTPTTKPELAPTTKPEAPTTPEELAPTTPEEPTPT 618
 Db 601 APTAPKPEAPTTKPEAPTTTKKLTPTTKPELAPTTKPEAPTTPEELAPTTPEEPTPT 660
 QY 619 PEAPPTPKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKGTAPTTPK 678
 Db 661 PEAPPTPKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKGTAPTTPKGTAPTTP 720
 QY 679 APTPKKAPKELAPTTTKESPTTSKAPTTKGTAPTTPKGTAPTTPKPEAPTTKPEAP 738
 Db 721 APTPKKAPKELAPTTTKESPTTSKAPTTKGTAPTTPKGTAPTTPKPEAPTTKPEAP 780
 QY 739 TAPTTKPEAPTTPKKAPKELAPTTTKGPTSTTSKAPTTKGTAPTTPKGTAPTTPKPEAP 798
 Db 781 TAPTTKPEAPTTPKKAPKELAPTTTKGPTSTTSKAPTTKGTAPTTPKGTAPTTPKPEAP 840
 QY 799 KPAPTPTPEPTTSVSTPTTKETPTTKESPTTSKAPTTKGTAPTTPKGTAPTTPKPEAP 858
 Db 841 KPAPTPTPEPTTSVSTPTTKETPTTKESPTTSKAPTTKGTAPTTPKGTAPTTPKPEAP 900
 QY 859 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPTAKMTKETATTTESKITATTTQV 918
 Db 901 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPTAKMTKETATTTESKITATTTQV 960
 QY 919 TSTTTQDTPPKITTLKTTTLAPKVTTKKTIITTEIMNKPBEFTAKPKDRATNSKATPK 978
 Db 961 TSTTTQDTPPKITTLKTTTLAPKVTTKKTIITTEIMNKPBEFTAKPKDRATNSKATPK 1020
 QY 979 PQKPTKAPKPTSTKPKTMRVRAKPTTTPRKMSTMPBLNPTSIAEAMLTITRPN 1038
 Db 1021 PQKPTKAPKPTSTKPKTMRVRAKPTTTPRKMSTMPBLNPTSIAEAMLTITRPN 1080
 QY 1039 QTPNSKLVEVNPKEBAGGAGETPHMLLRPHVFMPEVTPDMOVLPRVNPQGIINPMLS 1098
 Db 1091 QTPNSKLVEVNPKEBAGGAGETPHMLLRPHVFMPEVTPDMOVLPRVNPQGIINPMLS 1140

RESULT 8

ADK67912

ID ADK67912 standard; protein; 1311 AA.

XX AC ADK67912;

XX DT 06-MAY-2004 (first entry)

XX

DE Human extracellular messenger (EXMES) polypeptide.
 XX Human; extracellular messenger; EXMES: respiratory-gen.; anti-allergic;
 KW anti-inflammatory; anti-infective; antidiabetic; neuroprotective;
 KW muscular-gen.; antiarthritic; osteopathic; hepatotropic; antipsoriatic;
 KW virucide; fungicide; antiparasitic; protozoacide; antihelminthic;
 XX cytostatic; gene therapy.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..30 /label= Signal_peptide
 FT /notes "Spans residues 1 to 18, 20, 21, 24, 29 or 30
 according to identification method"
 XX WO2004013292-A2.
 XX 12-FEB-2004.
 XX 30-JUL-2003; 2003WO-US024084.
 XX 02-AUG-2002; 2002US-0400810P.
 PR 19-SEP-2002; 2002US-0412197P.
 PR 04-OCT-2002; 2002US-0416004P.
 PR 08-NOV-2002; 2002US-0424862P.
 XX (INCY-) INCYTE CORP.
 PA Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP,
 PI Richardson TW, Emerling BM, Lindquist EA, Chawia NK, Ramkumar J,
 PI Lee SJ;
 XX MPI; 2004-157116/15.
 DR N-PSDB; ADK67917.
 XX New extracellular messengers and nucleic acids, useful for diagnosing,
 PT treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes
 PT mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or
 PT autoimmune thyroiditis.
 PS Claim 60; SEQ ID NO 5; 165pp; English.
 XX The present sequence is that of novel human extracellular messenger
 CC (EXMES) Incyte ID NO: 7513018CD1 polypeptide. The protein shows homology
 CC to human megakaryocyte stimulating factor. The invention provides EXMES
 CC polynucleotides and polypeptides, as well as expression vectors, host
 CC cells, antibodies, agonists and antagonists, and methods for diagnosing,
 CC treating or preventing disorders associated with aberrant expression of
 CC EXMES, especially autoimmune and inflammatory disorders, cell
 CC proliferative disorders and endocrine disorders, e.g. adult respiratory
 CC distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's
 CC disease, diabetes mellitus, myasthenia gravis, osteoarthritis,
 CC osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal,
 CC parasitic, protozoal or helminthic infections, cancers, autoimmune
 CC thyroiditis, cretinism, Plummer's disease or thyroid carcinoma.
 CC Embodiments also provide methods for using the purified EXMES and/or
 CC their encoding polynucleotides for facilitating the drug discovery
 CC process, including determining of efficacy, dosage, toxicity and
 CC pharmacology, and for investigating the pathogenesis of diseases and
 CC medical conditions.
 XX Sequence 1311 AA;
 SQ Query Match 95.3%; Score 5607.9; DB 8; Length 1311;
 Best Local Similarity 95.4%; Pred. No. 6.1e-147;
 Matches 1047; Conservative 0; Mismatches 0; Indels 51; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFIQQVSSQDLSSCAGRGEGYSRDATCNCYNCQHYMECCPDF 60

Db 1 MAWKTLPIYLLLLSVFIQQVSSQDLSSCAGRGEGYSRDATCNCYNCQHYMECCPDF 60

QY 61 KEVCTAELSKRCRCEFSFERGECDDCAQCKYDKCCPDFSFCAEVHNPTSPSSKKAP 120

Db 61 KRVTAEILSCGRGCFESFERGECDDAQCKYDKCCPDYEFCAE----- 106
Qy 121 PPSGASQIKSTTKESPRPPNKKTKKVI ESEIEITEKVKDNKQRTKKKPTPKPPVDEA 180
Db 107 -----VKDNKQRTKKKPTPKPPVDEA 129
Qy 181 GSGLDNGDFKVTTPDTSTQHNKYSTSKITTAKEINPRPSLPNSDTSKETSITVNET 240
Db 130 GSGLDNGDFKVTTPDTSTQHNKYSTSKITTAKEINPRPSLPNSDTSKETSITVNET 189
Qy 241 TVETKETTNNKQTSDDGKEKTSKAKTOSIEKTSKADLAPTCKVLAKPTPKAETTTKGP 300
Db 190 TVETKETTNNKQTSDDGKEKTSKAKTOSIEKTSKADLAPTCKVLAKPTPKAETTTKGP 249
Qy 301 ALTTTKEPTPTPKEPASTTKEPTPTTIKSAPTTPKEPAPTPTTKSAPTTPKEPAPTPTK 360
Db 250 ALTTTKEPTPTPKEPASTTKEPTPTTIKSAPTTPKEPAPTPTTKSAPTTPKEPAPTPTK 309
Qy 361 EPAPTTKEPAPTPTTKAPATTKSAPTTPKEPAPTPTTKAPATTKSAPTTPKEPAPTPTK 420
Db 310 EPAPTTKEPAPTPTTKAPATTKSAPTTPKEPAPTPTTKAPATTKSAPTTPKEPAPTPTK 369
Qy 421 TPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP 480
Db 370 TPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP 429
Qy 481 APPTTKSAPTTTKEPAPTPTTKSAPTTPKEPAPTPTTKSAPTTPKEPAPTPTTK 540
Db 430 APPTTKSAPTTTKEPAPTPTTKSAPTTPKEPAPTPTTKSAPTTPKEPAPTPTTK 489
Qy 541 EPAPTTKEPAPTPTTKAPATTKSAPTTPKEPAPTPTTKAPATTKSAPTTPKEPAPTPTK 600
Db 490 EPAPTTKEPAPTPTTKAPATTKSAPTTPKEPAPTPTTKAPATTKSAPTTPKEPAPTPTK 549
Qy 601 TTPPELAPTTPEEPTPTTPEEAPATTPKAAAPNTPKAPATTTPKEPAPTTPKEPAPTTPKEP 660
Db 550 TTPPELAPTTPEEPTPTTPEEAPATTPKAAAPNTPKAPATTTPKEPAPTTPKEPAPTTPKEP 609
Qy 661 ETAPTTKGTAPTTLKEPAPTTPKAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTT 720
Db 610 ETAPTTKGTAPTTLKEPAPTTPKAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTT 669
Qy 721 KEAPATTKEPAPTTPKGTAPTTLKEPAPTTPKAPKELAPTTTKEPTSTSDKAPATT 780
Db 670 KEAPATTKEPAPTTPKGTAPTTLKEPAPTTPKAPKELAPTTTKEPTSTSDKAPATT 729
Qy 781 PKETAPTTKEPAPTTPKAPATTTPETPPPTSEVSTPTTKEPTTIHKSPDESTPELSA 840
Db 730 PKETAPTTKEPAPTTPKAPATTTPETPPPTSEVSTPTTKEPTTIHKSPDESTPELSA 789
Qy 841 EPTPKALENSPKEGVPPTTKPAATKPEMTTTAKDKTTERDLRTTPETTTAAAPKMTETA 900
Db 790 EPTPKALENSPKEGVPPTTKPAATKPEMTTTAKDKTTERDLRTTPETTTAAAPKMTETA 849
Qy 901 TTTTEKTESKITATTQVSTTTQDTTPPKITTLKTTTTLAPKVTTKTITTEIMNKE 960
Db 850 TTTTEKTESKITATTQVSTTTQDTTPPKITTLKTTTTLAPKVTTKTITTEIMNKE 909
Qy 961 ETAPKDRATNSKATTPKOKPTKAPKKTPTS TKKBTMPKVRKPTTTPKMTSTMPEL 1020
Db 910 ETAPKDRATNSKATTPKOKPTKAPKKTPTS TKKBTMPKVRKPTTTPKMTSTMPEL 969
Qy 1021 NPTSRIAEAMLOTTTRNQTPNSKLVENPKSEDAAGAGETPHMLLRPHVFPEVTPDM 1080
Db 970 NPTSRIAEAMLOTTTRNQTPNSKLVENPKSEDAAGAGETPHMLLRPHVFPEVTPDM 1029
Qy 1081 DYLPRVNOGIIINPMLS 1098
Db 1030 DYLPRVNOGIIINPMLS 1047

RESULT 9

ADK65819
ID ADK65819 standard; protein; 1320 AA.
XX AC
XX ADK65819;
XX AC
DT 06-MAY-2004 (first entry)
XX DE
XX Angiogenesis-differentially expressed protein ANH0316.
XX DE
XX cyotostatic; cardiant; vasotropic; antiarteriosclerotic;
KW angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
KW gene expression; cancer; coronary artery disease; myocardial ischemia;
KW coronary arteriosclerosis; forensic medicine.
XX OS
XX Homo sapiens.
XX PN W02003066831-A2.
XX PD 14-AUG-2003.
XX PF 07-FEB-2003; 2003WO-US003848.
XX PR 07-FEB-2002; 2002US-00067482.
XX PR 10-JUN-2002; 2002US-00164595.
XX PR 16-AUG-2002; 2002US-0403649P.
XX PR 03-JAN-2003; 2003US-0437746P.
XX PA (ORIG-) ORIGENE TECHNOLOGIES INC.
XX PI Sun Z, Li X, Kovacs KF, Fan W, Jay G;
XX DR WPI; 2003-731502/69.
XX DR N-PSDB; ADK65818.
XX PT Determining the angiogenic index of a tissue or cell sample using
PT expression levels of differentially expressed genes, useful for
PT diagnosing or treating cancer, coronary artery disease, myocardial
PT ischemia and/or arteriosclerosis.
XX PS Claim 23; SEQ ID NO 58; 296pp; English.
XX CC The invention relates to a method of determining the angiogenic index of
CC a tissue or cell sample comprising assessing, in a sample, the expression
CC levels of one or more differentially-expressed gene from any of 34 DNA
CC sequences, given in the specification, where the levels are indicative of
CC the angiogenic index. The methods and compositions of the present
CC invention are useful for diagnosing, preventing and/or treating cancer,
CC coronary artery disease, myocardial ischemia or coronary
CC arteriosclerosis. They can also be used in research, drug discovery and
CC forensic medicine involving angiogenesis. This sequence corresponds to
CC one of the differentially expressed proteins of the invention.
XX SQ Sequence 1320 AA;
Query Match 95.1%; Score 5598.8; DB 7; Length 1320;
Best Local Similarity 96.0%; Pred. No. 1.1e-146;
Matches 1054; Conservative 0; Mismatches 2; Indels 42; Gaps 2;
Qy 1 MAWKTLPIYLILLLSVFVIQQVSSQDLSSCAGRGEGYSRDTATCNDYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLILLLSVFVIQQVSSQ----- 25
Qy 61 KRVTAEILSCGRGCFESFERGECDDAQCKYDKCCPDYEFCAE VHNPTSPSSKKAP 120
Db 26 -----ELUSCKGRGCFESFERGECDDAQCKYDKCCPDYEFCAE VHNPTSPSSKKAP 79
Qy 121 PPSGASQIKSTTKESPRPPNKKTKKVI ESEIEITEKVKDNKQRTKKKPTPKPPVDEA 180
Db 80 PPSGASQIKSTTKESPRPPNKKTKKVI ESEIEITE-VKDNKQRTKKKPTPKPPVDEA 138
Qy 181 GSGLDNGDFKVTTPDTSTQHNKYSTSKITTAKEINPRPSLPNSDTSKETSITVNET 240
Db 139 GSGLDNGDFKVTTPDTSTQHNKYSTSKITTAKEINPRPSLPNSDTSKETSITVNET 198

QY 241 TVETKETTTNNKQTSIDGKEKTTSAKETQSIEXTSAKDLAPTSKVLAKPTPKAETTTKGP 300
 Db 199 TVETKETTTNNKQTSIDGKEKTTSAKETQSIEXTSAKDLAPTSKVLAKPTPKAETTTKGP 258
 QY 301 ALTPKPEPTTTKBPASATTPKPEPTTTIKSAPTTKBPAPTTKSAPTTKBPAPTTTK 360
 Db 259 ALTPKPEPTTTKBPASATTPKPEPTTTIKSAPTTKBPAPTTKSAPTTKBPAPTTTK 318
 QY 361 EPAPTTTKBPAPTTTKBPAPTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 420
 Db 319 EPAPTTTKBPAPTTTKBPAPTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 378
 QY 421 TPKEPAPTTKBPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 480
 Db 379 TPKEPAPTTKBPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 438
 QY 481 APPTTKSAPTTTKBPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 540
 Db 439 APPTTKSAPTTTKBPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 498
 QY 541 EPAPTTTKBPAPTTTKBPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 600
 Db 499 EPAPTTTKBPAPTTTKBPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 558
 QY 601 TPPEELAPTTPEPTTTPEEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 660
 Db 559 TPPEELAPTTPEPTTTPEEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 618
 QY 661 ETAPTTTKBPAPTTTKBPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 720
 Db 619 ETAPTTTKBPAPTTTKBPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 678
 QY 721 KEAPTTTKBPAPTTTKBPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 780
 Db 679 KEAPTTTKBPAPTTTKBPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 738
 QY 781 PKETAPTTTKBPAPTTTKBPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 840
 Db 739 PKETAPTTTKBPAPTTTKBPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 798
 QY 841 EPTPKALENSPKBPAPTTTKBPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 900
 Db 799 EPTPKALENSPKBPAPTTTKBPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKBPAPTT 858
 QY 901 TTTEKTTESKITATTTQVTSITTTQDTPPKKLTTLTKTTTLAKVTTTKKTTITTTIMNKP 960
 Db 859 TTTEKTTESKITATTTQVTSITTTQDTPPKKLTTLTKTTTLAKVTTTKKTTITTTIMNKP 918
 QY 961 ETAPKPKDRATNSKATTPPKQPKTAPKAPKPTSTKPKMTMPVRKPKTTPPKKMTSTMP 1020
 Db 919 ETAPKPKDRATNSKATTPPKQPKTAPKAPKPTSTKPKMTMPVRKPKTTPPKKMTSTMP 978
 QY 1021 NPTSEIAEAMLTQTTTRPNQNSKLVNPKSEBAGAGBETPHMLLRPHVFMPEVTPDM 1080
 Db 979 NPTSEIAEAMLTQTTTRPNQNSKLVNPKSEBAGAGBETPHMLLRPHVFMPEVTPDM 1038
 QY 1081 DYLPRVFNQGIINPMLS 1098
 Db 1039 DYLPRVFNQGIINPMLS 1056

RESULT 10

ADK67911

ID ADK67911 standard; protein; 1270 AA.

XX AC ADK67911;

XX DT 06-MAY-2004 (first entry)

XX DE Human extracellular messenger (EXMES); polypeptide.

XX KW Human; extracellular messenger; EXMES; respiratory-gen.; antiallergic;

KW

KW

KW

KW

XX

OS

XX

FH

FT

FT

FT

FT

XX

PN

PD

XX

PF

XX

PR

PR

PR

PR

XX

FA

XX

PI

PI

PI

XX

DR

DR

XX

PT

PT

PT

PT

XX

FS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Query Match

Best Local Similarity 90.8%; Score 5343.8; DB 8; Length 1270;

Matches 1006; Conservative 0; Mismatches 0; Indels 92; Gaps 2;

QY 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSRDRATNCNDYNOHYNECCPDF 60
 |||||
 Db 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSRDRATNCNDYNOHYNECCPDF 25

QY 61 KRVCCTAELSKGRCFESFERGECDCDAQCKYDKCCPDYEFCAEVHNPTSPSSKKAP 120
 |||||
 Db 26 -----ELSKGRCFESFERGECDCDAQCKYDKCCPDYEFCAE----- 65

antiasthmatic; antiinflammatory; antidiabetic; neuroprotective;
 muscular-gen.; antiarthritic; osteopathic; hepatotropic; antiparasitic;
 virucide; fungicide; antiparasitic; protozoacide; antihelminthic;
 cytostatic; gene therapy.

Homo sapiens.

Key Location/Qualifiers

Peptide 1..29

/label= Signal_peptide

/note= "Spans residues 1 to 18, 20, 21, 24, 27 or 29 according to identification method"

WO2004013292-A2.

12-FEB-2004.

30-JUL-2003; 2003WO-US024084.

02-AUG-2002; 2002US-0400810P.

19-SEP-2002; 2002US-0412197P.

04-OCT-2002; 2002US-0416004P.

08-NOV-2002; 2002US-0424862P.

(INCY-) INCYTE CORP.

Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP;

Richardson TW, Emerling BM, Lindquist EA, Chawla NK, Ramkumar J;

Lee SY;

WPI; 2004-157116/15.

N-PSDB; ADK67916.

New extracellular messengers and nucleic acids, useful for diagnosing,
 treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes
 mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or
 autoimmune thyroiditis.

Claim 59; SEQ ID NO 4; 165pp; English.

The present sequence is that of novel human extracellular messenger
 (EXMES) Incyte ID NO: 7513017CD1 polypeptide. The invention shows homology
 to human megakaryocyte stimulating factor. The invention provides EXMES
 polynucleotides and polypeptides, as well as expression vectors, host
 cells, antibodies, agonists and antagonists, and methods for diagnosing,
 treating or preventing disorders associated with aberrant expression of
 EXMES, especially autoimmune and inflammatory disorders, cell
 proliferative disorders and endocrine disorders, e.g. adult respiratory
 distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's
 disease, diabetes mellitus, myasthenia gravis, osteoarthritis, Crohn's
 disease, osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal,
 parasitic, protozoal or helminthic infections, cancers, autoimmune
 thyroiditis, cretinism, Plummer's disease or thyroid carcinoma.
 Embodiments also provide methods for using the purified EXMES and/or
 their encoding polynucleotides for facilitating the drug discovery
 process, including determining of efficacy, dosage, toxicity and
 pharmacology, and for investigating the pathogenesis of diseases and
 medical conditions.

Sequence 1270 AA;

CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats
 CC of a motif having at least 50% identity to the sequence KEPAPTT
 CC (AAB29774). The invention also relates to a nucleic acid encoding a human
 CC MSF-derived tribonection; a biocompatible composition comprising a human
 CC tribonection for inhibiting tissue adhesion formation; and a method of
 CC diagnosing osteoarthritis or a predisposition to osteoarthritis by
 CC measuring the amount of MSF or its fragment in a biological sample of a
 CC mammal, wherein an increased amount of MSF compared to a control
 CC indicates the presence of or predisposition to developing osteoarthritis.
 CC The tribonection and DNA encoding it are useful in the treatment of
 CC osteoarthritis, where they may be used for lubricating mammalian joints,
 CC such as articulating joints of humans, dogs or horses. The tribonection,
 CC when formulated as a membrane, foam, gel or fibre, is useful for
 CC inhibiting adhesion between two surfaces such as the injured tissues of a
 CC mammal, where the injury is caused by a surgical insertion or trauma, or
 CC an artificial device e.g., an orthopaedic implant. In particular, one of
 CC the surfaces is pericardial tissue. DNA encoding a tribonection may be
 CC used in gene therapy. The present sequence represents a substantial
 CC portion of a human MSF-derived tribonection
 CC
 XX
 SQ Sequence 902 AA;
 Query Match 48.9%; Score 2880.1; DB 4; Length 902;
 Best Local Similarity 71.4%; Pred. No. 9.9e-72;
 Matches 675; Conservative 27; Mismatches 74; Indels 169; Gaps 66;
 QY 158 VKDNKKNRTKKPTKPPVVDAGSGLDNGDFKVTTPDSTTQHKNKSVSPKITTAKPIN 217
 DB 1 VKDNKKNRTKKPTKPPVVDAGSGLDNGDFKVTTPDSTTQHKNKSVSPKITTAKPIN 60
 QY 218 PRPSLPNSDTSKETSLSLVNKKETTVETKTTNNKQTSIDGKEKTSIAKETOSIEKTSK 277
 DB 61 PRPSLPNSDTSKETSLSLVNKKETTVETKTTNNKQTSIDGKEKTSIAKETOSIEKTSK 120
 QY 278 DLAPTSKVLAKPTPKAETTTKGPALTTPKETPTTPKEPASTTPKEPTTTKSAPTTPK 337
 DB 121 DLAPTSKVLAKPTPKAETTTKGPAL-----TTPKEPASTTPKEPTTTKSAPTTPK 172
 QY 338 EPAPTTTKSAPTTPKEPAPTTKBPAPTTKEBPATTTKSAPTTPKBPAPTT 397
 DB 173 EPAPTTTKSAPTTPKEPAPTTKBPAPTTKEBPATTTKSAPTTPKBPAPTT 228
 QY 398 PKKPAPTTKEBPATTPKEPTTPTPKEBPATTTKEBPATTPKBPAPTPAKKBPATTPKEPA 457
 DB 229 -KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTTKEBPATTT-KEBPATTT-KEPA 281
 QY 458 PTTTKBPAPTTTKBPAPTTKEBPATTTKSAPTTTKSAPTTTKSAPTTPK 517
 DB 282 PTT-KEPAP-TTKBPAPTT-KEPAPTT-KEPAP-TTKBPAPTTKEBPATTT-KEPAP-TTKE 335
 QY 518 PAPTTPKEBPATTPKBPAPTTKEBPATTTKEBPATTTTKKBPAPTAPKEPAPTTKEAPT 577
 DB 336 PAPTTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAP-TTKEPAPTT-KEPAPTT-KEPAPT 388
 QY 578 TPKKLPTTPPEKLAPTTPKEBPATTPPELAPTTPEBPPTTTPPEBPAPTTKAAAPTPKE 637
 DB 389 T-KEPAPTTKEP-APTTPKE-PAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KE 440
 QY 638 PAPTTPKEBPATTPKBPAPTTKETAPTTPKGAPTTLKBPAPTTKBPAP--KELAPTT 695
 DB 441 PAPTTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTTKEPAP-T 493
 QY 696 TKBPSTTSDDKAPTTPKGAPTTTPKEBPATTTKEBPATTTKGATPTLKBPAPTTKBP 755
 DB 494 TKBPAPTTKE-PAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEP 546
 QY 756 AP--KELAPTTTKGPTSTTSDDKAPTTPKETAPTTKEBPAPTTKBPAPTTTPPTPTTS 813
 DB 547 APPTTKBPAP-TTKBPAPTTKE-PAPTT-KEPAPTT-KEPAPTT-KEPAPTTKE-PAPTTK 600
 QY 814 EVSTPTTTKBPPTTHKSPDSESTPBLSAEPTPKALENSPKPEPGVPTTKTAPATKPEMTT 873
 DB 601 EPA--PTTKBPAPTTKEPAPTT---KEPAP-----TTKEP-APTTPKEPAPT----- 640

QY 874 KKYTERDLRTTPETTAAPKMTKETATTEKTESKITATTQVTTSTTQDTPFKIIT 933
 DB 641 -----TKEPAPTT-----KEPAPTTKEPAPTTKEPAP----- 667
 QY 934 LKTTTLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATTPKPKQKTPKAPKKTSTK 993
 DB 668 -----TTKEPAPTTKEP-APT-TK 684
 QY 994 KPTMPRVKPKTPTTPRKQKSTMPBELNPTSRISAEAMLOTTFPNTQNSKLVENPKSE 1053
 DB 685 EP-----APTPTPRKQKSTMPBELNPTSRISAEAML-TTTPNQTNSKLVENPKSE 735
 QY 1054 DAGGAEGETHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1098
 DB 736 DAGGAEGETHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 780
 RESULT 14
 ABUS3254
 ID ABUS3254 standard; protein; 513 AA.
 AC ABUS3254;
 XX
 DT 14-APR-2003 (first entry)
 XX
 DE Human testes-derived DKFZphtes3_4019 homologue #3.
 XX
 KW Human; gene therapy; vaccine; disease treatment; detection.
 XX
 OS Homo sapiens.
 XX
 PN WO200112659-A2.
 XX
 PD 22-FEB-2001.
 XX
 PF 18-AUG-2000; 2000WO-IB001496.
 XX
 PR 18-AUG-1999; 99US-0149499P.
 XX
 PR 28-SEP-1999; 99US-0156503P.
 XX
 PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
 XX
 PI Wiemann S;
 XX
 DR WPI; 2001-327840/34.
 XX
 PT Nucleic acids having the sequences of clones isolated from libraries of
 PT different human tissues, useful in recombinant DNA methodologies.
 XX
 PS Example III; Page 893; 1095pp; English.
 XX
 CC This invention describes novel polynucleotides and polypeptides isolated
 CC from human cDNA libraries which can be used for gene therapy or in
 CC vaccines. The polynucleotides of the invention and antibodies encoded by
 CC them may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate polypeptide expression. The products of the
 CC invention may also be used to identify modulators of expression and
 CC activity and to down regulate expression and activity. The antibodies of
 CC the invention may also be used as diagnostic agents for detecting the
 CC presence of polypeptides in samples. This sequence represents a homologue
 CC of a polypeptide described in the disclosure of the invention
 XX
 SQ Sequence 513 AA;
 Query Match 46.8%; Score 2757; DB 4; Length 513;
 Best Local Similarity 100.0%; Pred. No. 1.2e-68;
 Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 165 RTKKKPTKPPVVDAGSGLDNGDFKVTTPDSTTQHKNKSVSPKITTAKPINRPSLPP 224
 DB 1 RTKKKPTKPPVVDAGSGLDNGDFKVTTPDSTTQHKNKSVSPKITTAKPINRPSLPP 60

QY 225 NSDTSKETSIVNKKETTTTNTKQSTGDKETTSKETSIAKDLAPTSK 284
 DB 61 NSDTSKETSIVNKKETTTTNTKQSTGDKETTSKETSIAKDLAPTSK 120
 QY 285 VLAKPTPKATTTKGPALITPKPTTTPKEPASTTPKEPTTTIKSAPTTKEDAPTTT 344
 DB 121 VLAKPTPKATTTKGPALITPKPTTTPKEPASTTPKEPTTTIKSAPTTKEDAPTTT 180
 QY 345 KSAPTTPKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 404
 DB 181 KSAPTTPKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 240
 QY 405 TPKEPAPTTPKPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 464
 DB 241 TPKEPAPTTPKPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 300
 QY 465 APTTTKEPPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 524
 DB 301 APTTTKEPPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 360
 QY 525 EPAPTTPKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 584
 DB 361 EPAPTTPKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 420
 QY 585 TTPEKLAPTTTPKEPAPTTTPEELAPTTPEPTTTPKEPAPTTTPEEAPTTTPEEAPTTTPEEAPTTT 644
 DB 421 TTPEKLAPTTTPKEPAPTTTPEELAPTTPEPTTTPKEPAPTTTPEEAPTTTPEEAPTTTPEEAPTTT 480
 QY 645 EPAPTTPKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 677
 DB 481 EPAPTTPKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 513

RESULT 15

AAR80041
 ID AAR80041 standard; protein; 452 AA.
 XX
 AC AAR80041;
 XX
 DT 25-MAR-2003 (revised)
 DT 10-APR-1996 (first entry)
 XX
 DE Human megakaryocytopoietin protein.
 XX
 KW Human; megakaryocytopoietin; wheat germ agglutinin; heparin;
 KW megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;
 KW multipotential stem cell.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 393..396
 FT /note= "unspecified amino acids"
 FT Misc-difference 444..446
 FT /note= "unspecified amino acids"
 XX
 PN W09523861-A1.
 XX
 PD 08-SEP-1995.
 XX
 PF 06-MAR-1995; 95WO-CN000015.
 XX
 PR 04-MAR-1994; 94CN-00112066.
 XX
 PA (SHAN-) SHANGHAI BEITE BIOTECHNOLOGY CO LTD.
 XX
 PI Gu X, Han Z, Shen Q;
 XX
 DR WPI; 1995-320576/41.
 DR N-FSD; AAT04546.
 XX
 FT New haematopoietic cell growth factor - used for treating
 FT thrombocytopenia and hematocytopenia.

XX

Example; Page 23; 36pp; Chinese.

XX This sequence represents the human megakaryocytopoietin (MPO) protein.
 CC This sequence was purified using a carrier which can couple wheat germ
 CC agglutinin and heparin to separate MPO. Fragments of this sequence (see
 CC AAR80039 and AAR80040) were used to produce the amplification primers
 CC shown in AAT04544 and AAT04545. The fragments amplified by these primers
 CC can then be used as probes to screen human cDNA libraries for MPO cDNA.
 CC The MPO cDNA can then be inserted into a plasmid which is used to
 CC transform cells to produce MPO. The MPO sequence is capable of promoting
 CC colony formation of megakaryocytes, enlarging the size of megakaryocytes
 CC and stimulating the proliferation of multipotential stem cells. The
 CC factor may be used for treating thrombocytopenia and hematocytopenia. The
 CC purification method can be used to isolate MPO from human urine or serum
 CC of patients with aplastic anaemia, and from animal blood or urine by
 CC radiation exposing the animals to induce aplastic anaemia. (Updated on 25
 CC -MAR-2003 to correct PA field.)
 XX

SQ Sequence 452 AA;

Query Match 37.8%; Score 2222.8; DB 2; Length 452;
 Best Local Similarity 92.1%; Pred. No. 6.1e-54;
 Matches 419; Conservative 3; Mismatches 21; Indels 12; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
 DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
 QY 61 KRVTAEISCKGRCPESFERGECDDAQQKYDKCCPDYSEFCABVHNPTSPSSKKAP 120
 DB 61 KRVTAEISCKGRCPESFERGECDDAQQKYDKCCPDYSEFCABVHNPTSPSSKKAP 120
 QY 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIESBEITEKVKDNKNRTKKKPTPKPPVDEA 180
 DB 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIESBEITE-KVDNKNRTKKKPTPKPPVDEA 179
 QY 181 GSGLDNGDFKVTTPDTSTTQHNKVTSPKITTAKPINRPSPSPNSDTSKETSIVNKKET 240
 DB 180 GSGLDNGDFKVTTPDTSTTQHNKVTSPKITTAKPINRPSPSPNSDTSKETSIVNKKET 239
 QY 241 TVETKETTNTKQSTGDKETTSKETSIAKDLAPTSKVLAKPTPKAETTTKGP 300
 DB 240 TVETKETTNTKQSTGDKETTSKETSIAKDLAPTSKVLAKPTPKAETTTKGP 299
 QY 301 ALTTPEPTTTPKEPASTTPKEPTTTIKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 360
 DB 300 ALTTPEPTTTPKEPASTTPKEPTTTIKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 359
 QY 361 EPAPTTPKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 420
 DB 360 EPAPTTPKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 408
 QY 421 TPKEPAPTTKEPAPTTTPKEPAPTAAPKAPAPTTTKE 455
 DB 409 LPRSLHPTTKEPAPTTTPKEPAPTAAPKAPAPLPPL 443

Search completed: October 13, 2004, 11:36:52
 Job time : 94.5238 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 19.0865 Seconds
(without alignments)
3815.116 Million cell updates/sec

Title: SEQ1-C
Perfect score: 587
Sequence: 1 NAWKTLPIYLLLSVFIQ.....DMDYLRVFNQGIINPMLS 1098

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/prodata/1/aa/5A-COMB.pep: *
2: /cgn2_6/prodata/1/aa/5B-COMB.pep: *
3: /cgn2_6/prodata/1/aa/6A-COMB.pep: *
4: /cgn2_6/prodata/1/aa/6B-COMB.pep: *
5: /cgn2_6/prodata/1/aa/PCTUS-COMB.pep: *
6: /cgn2_6/prodata/1/aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5872.8	99.8	1140	4	US-07-757-022B-104
2	5872.8	99.8	1404	4	US-07-757-022B-2
3	5872.8	99.8	1404	4	US-07-757-022B-62
4	5872.8	99.8	1404	4	US-09-298-970A-1
5	5871.9	99.7	1361	4	US-07-757-022B-40
6	5863.8	99.6	1404	4	US-10-164-595-78
7	5847.2	95.9	1364	4	US-07-757-022B-48
8	5808.7	95.3	1363	4	US-07-757-022B-52
9	5607.9	95.3	1311	4	US-07-757-022B-42
10	5607.8	95.3	1320	4	US-07-757-022B-46
11	5607.8	95.3	1320	4	US-07-757-022B-60
12	5598.8	95.1	1320	4	US-10-164-595-58
13	5466.9	92.9	1022	4	US-07-757-022B-84
14	5399.2	91.7	1314	4	US-07-757-022B-50
15	5383.1	91.4	1049	4	US-07-757-022B-58
16	5383.1	91.4	1313	4	US-07-757-022B-142
17	5343.8	90.8	1038	4	US-07-757-022B-74
18	5343.8	90.8	1270	4	US-07-757-022B-44
19	5011	85.1	941	4	US-07-757-022B-14
20	1396.9	23.7	5179	4	US-09-538-093-1258
21	1132.9	19.2	207	4	US-07-757-022B-116
22	1132.9	19.2	207	4	US-07-757-022B-136
23	1127.3	19.1	8991	4	US-08-714-741-32
24	975.8	16.6	220	4	US-07-757-022B-96
25	926.6	15.7	463	4	US-07-757-022B-54
26	913.8	15.5	209	4	US-07-757-022B-94
27	908.8	15.4	208	4	US-07-757-022B-132

28 905.4 15.4 296 4 US-07-757-022B-70
29 899.7 15.3 231 4 US-07-757-022B-30
30 893 15.2 204 4 US-07-757-022B-92
31 886.8 15.1 192 4 US-07-757-022B-90
32 876.8 14.9 172 4 US-07-757-022B-88
33 872.5 14.8 188 4 US-07-757-022B-32
34 871 14.8 156 4 US-07-757-022B-106
35 871 14.8 156 4 US-07-757-022B-118
36 868.9 14.8 157 4 US-07-757-022B-102
37 868.9 14.8 157 4 US-07-757-022B-114
38 847.3 14.4 3256 4 US-09-919-172-98
39 847.3 14.4 3256 4 US-09-976-594-22
40 847.3 14.4 3256 4 US-09-919-039-21
41 845.9 13.9 3118 3 US-08-579-181-1
42 807.2 13.7 2972 3 US-09-579-181-2
43 800 13.6 141 4 US-07-757-022B-80
44 778 13.2 2142 4 US-09-538-092-1142
45 767.9 13.0 1837 3 US-08-928-361B-5

ALIGNMENTS

RESULT 1
US-07-757-022B-104
; Sequence 104, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: AMINO ACID

Sequence 70, Appl
Sequence 30, Appl
Sequence 92, Appl
Sequence 90, Appl
Sequence 88, Appl
Sequence 32, Appl
Sequence 106, App
Sequence 118, App
Sequence 102, App
Sequence 114, App
Sequence 98, Appl
Sequence 21, Appl
Sequence 22, Appl
Sequence 1, Appl
Sequence 80, Appl
Sequence 1142, Ap
Sequence 5, Appl

QY 61 KRVCTAELSKGRGCFESFERGECDDAQQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
DB 61 KRVCTAELSKGRGCFESFERGECDDAQQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY 121 PPGASQTIKSTTKRSPKPNKKTKKVISESEITE----- 156
DB 121 PPGASQTIKSTTKRSPKPNKKTKKVISESEITE----- 180
QY 157 -----KVNDKKNRTKKKTKKVISESEITE----- 198
DB 157 -----KVNDKKNRTKKKTKKVISESEITE----- 240
QY 181 KIKSKNSAANRELQKLLKVNDKKNRTKKKTKKVISESEITE----- 258
DB 181 KIKSKNSAANRELQKLLKVNDKKNRTKKKTKKVISESEITE----- 300
QY 199 TOHNVSTSPKITTAKPINRPSLPDSGTSKETSITVKNKETTIVETKETTINKOTSDG 258
DB 241 TOHNVSTSPKITTAKPINRPSLPDSGTSKETSITVKNKETTIVETKETTINKOTSDG 300
QY 259 KEKTSKETSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPAS 318
DB 301 KEKTSKETSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPAS 360
QY 319 TTPKEPTTTIKSAPTTKBPAPTTKGAAPTTKBPAPTTTKEBPAPTTKBPAPTTTKEP 378
DB 361 TTPKEPTTTIKSAPTTKBPAPTTKGAAPTTKBPAPTTTKEBPAPTTTKEP 420
QY 379 APTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKEBPAPTTTKEP 438
DB 421 APTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKEBPAPTTTKEP 480
QY 439 EPAPTAPKKPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPT 498
DB 481 EPAPTAPKKPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPT 540
QY 499 TTKSAPTTKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEP 558
DB 541 TTKSAPTTKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEP 600
QY 559 APTAKEPAPTTKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPT 618
DB 601 APTAKEPAPTTKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPT 660
QY 619 PEPAPTTKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEP 678
DB 661 PEPAPTTKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEP 720
QY 679 APTTKKAPKELAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEP 738
DB 721 APTTKKAPKELAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEP 780
QY 739 TAPTTKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEP 798
DB 781 TAPTTKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEP 840
QY 799 KPAPTTPPTTSEVSTPTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEP 858
DB 841 KPAPTTPPTTSEVSTPTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEP 900
QY 859 TKTPAATKEMITAKDKTERDLRTTPTTAAKPMKETAATTTKETSITATTTQV 918
DB 901 TKTPAATKEMITAKDKTERDLRTTPTTAAKPMKETAATTTKETSITATTTQV 960
QY 919 TSTTTQDTPFKITLTKTTLAPKVTTKKTTTTEIMNKPEETAKPDRAINSKATTPK 978
DB 961 TSTTTQDTPFKITLTKTTLAPKVTTKKTTTTEIMNKPEETAKPDRAINSKATTPK 1020
QY 979 POKPTKAPKPTSTKPKPMRVRKPTTTPRKMSTWMPNLNPSRIAEAMLQTTTRPN 1038
DB 1021 POKPTKAPKPTSTKPKPMRVRKPTTTPRKMSTWMPNLNPSRIAEAMLQTTTRPN 1080
QY 1039 QTPNSKLVENPKSADGAEGETPHMLLRPHVFPVEVTPDMDYLRVFNQGIINPMLS 1098
DB 1081 QTPNSKLVENPKSADGAEGETPHMLLRPHVFPVEVTPDMDYLRVFNQGIINPMLS 1140

RESULT 3
US-07-757-022B-62
; Sequence 62, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserk, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-62

Query Match 99.8%; Score 5872.8; DB 4; Length 1404;
Best Local Similarity 96.3%; Pred. No. 5.5e-171;
Matches 1098; Conservative 0; Mismatches 0; Indels 42; Gaps 1;
QY 1 MAWTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCYDNCQHYMECCPDF 60
DB 1 MAWTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCYDNCQHYMECCPDF 60
QY 61 KRVCTAELSKGRGCFESFERGECDDAQQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
DB 61 KRVCTAELSKGRGCFESFERGECDDAQQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY 121 PPGASQTIKSTTKRSPKPNKKTKKVISESEITE----- 156
DB 121 PPGASQTIKSTTKRSPKPNKKTKKVISESEITE----- 180
QY 157 -----KVNDKKNRTKKKTKKVISESEITE----- 198

```
Db 181 KIKSSKNSAANRELOKLVKNDKNKRTKKKPTKPPVVVDAGSLDNGDFKVTTPDST 240
QY 199 TQHNKYSTSPKLTITAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTNNKOTSDG 258
Db 241 TQHNKYSTSPKLTITAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTNNKOTSDG 300
QY 259 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPEKPTTTPKEPAS 318
Db 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPEKPTTTPKEPAS 360
QY 319 TTPKEPTPTTIKSAPTTKEPAPTTTKSAPTTKEPAPTTTKSAPTTKEPAPTTTKKEP 378
Db 361 TTPKEPTPTTIKSAPTTKEPAPTTTKSAPTTKEPAPTTTKSAPTTKEPAPTTTKKEP 420
QY 379 APTTTKSAPTTTPKEPAPTTTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 438
Db 421 APTTTKSAPTTTPKEPAPTTTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 480
QY 439 EPAPTAPKPAPTTTPKEPAPTTTKSAPTTTKESPTTTPKEPAPTTTKSAPTTTKKEP 498
Db 481 EPAPTAPKPAPTTTPKEPAPTTTKSAPTTTKESPTTTPKEPAPTTTKSAPTTTKKEP 540
QY 499 TTKSAPTTTPKEPAPTTTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 558
Db 541 TTKSAPTTTPKEPAPTTTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 600
QY 559 APTAPKPAPTTTPKEPAPTTTKSAPTTTKESPTTTPKEPAPTTTKSAPTTTKKEP 618
Db 601 APTAPKPAPTTTPKEPAPTTTKSAPTTTKESPTTTPKEPAPTTTKSAPTTTKKEP 660
QY 619 PEAPAPTTTPKAAAPNTPEKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 678
Db 661 PEAPAPTTTPKAAAPNTPEKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 720
QY 679 APTTPKPAKELAPTTTKETSTTSKAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTPK 738
Db 721 APTTPKPAKELAPTTTKETSTTSKAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTPK 780
QY 739 TAPTTLKPAPTTTPKPAKELAPTTTKGTSTTSKAPTTTPKGTAPTTTPKEPAPTTPK 798
Db 781 TAPTTLKPAPTTTPKPAKELAPTTTKGTSTTSKAPTTTPKGTAPTTTPKEPAPTTPK 840
QY 799 KPAPTTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAETTPKALENSKPEGVPT 858
Db 841 KPAPTTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAETTPKALENSKPEGVPT 900
QY 859 TKTAAATKPEMTTAKDKTTERDLRTPETTTAAPKMTKETATTTKETSKITATTQV 918
Db 901 TKTAAATKPEMTTAKDKTTERDLRTPETTTAAPKMTKETATTTKETSKITATTQV 960
QY 919 TSTTTQDTPPKITTLTKTLLAPKVTTTKKTTTTEIMNKPETAKPKDRATNSKATTPK 978
Db 961 TSTTTQDTPPKITTLTKTLLAPKVTTTKKTTTTEIMNKPETAKPKDRATNSKATTPK 1020
QY 979 POKTAPKPKTSTKPKTMRVVKPKTTPTRKMTSTMPELAPTSIAEAMLOTTTRN 1038
Db 1021 POKTAPKPKTSTKPKTMRVVKPKTTPTRKMTSTMPELAPTSIAEAMLOTTTRN 1080
QY 1039 QTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMYLPVFNQGIINPMLS 1098
Db 1081 QTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMYLPVFNQGIINPMLS 1140
```

RESULT 4

```
US-09-298-970A-1
; Sequence 1, Application US/09298970A
; Patent No. 6743774
; GENERAL INFORMATION:
; APPLICANT: Jay, Gregory D.
; TITLE OF INVENTION: TRIBONECTINS
; FILE REFERENCE: 21486-026
; CURRENT APPLICATION NUMBER: US/09/298,970A
; CURRENT FILING DATE: 2001-06-19
```

```
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-298-970A-1

Query Match          99.8%; Score 5872.8; DB 4; Length 1404;
Best Local Similarity 96.3%; Pred. No. 5.5e-171;
Matches 1098; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MAWKTLPIYLLLLLSVFVIQVSSQDLSSCAGCGEGYSRDATCNCYNQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLLSVFVIQVSSQDLSSCAGCGEGYSRDATCNCYNQHYMECCPDF 60
QY 61 KRVCYTAELCKGRCFESPERGECDDAQCKKYDKCCPDYESFCAYVHNPTSPSSKAP 120
Db 61 KRVCYTAELCKGRCFESPERGECDDAQCKKYDKCCPDYESFCAYVHNPTSPSSKAP 120
QY 121 PPSGASQTIKSTTKRSPPNKKTKKVIESBEITE----- 156
Db 121 PPSGASQTIKSTTKRSPPNKKTKKVIESBEITE----- 156
QY 157 -----KVKONKQNRKYKKPTPKPPVVVDAGSLDNGDFKVTTPDST 198
Db 181 KIKSSKNSAANRELOKLVKNDKNKRTKKKPTKPPVVVDAGSLDNGDFKVTTPDST 240
QY 199 TQHNKYSTSPKLTITAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTNNKOTSDG 258
Db 241 TQHNKYSTSPKLTITAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTNNKOTSDG 300
QY 259 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPEKPTTTPKEPAS 318
Db 301 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPEKPTTTPKEPAS 360
QY 319 TTPKEPTPTTIKSAPTTKEPAPTTTKSAPTTTKESPTTTPKEPAPTTTKSAPTTTKKEP 378
Db 361 TTPKEPTPTTIKSAPTTKEPAPTTTKSAPTTTKESPTTTPKEPAPTTTKSAPTTTKKEP 420
QY 379 APTTTKSAPTTTPKEPAPTTTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 438
Db 421 APTTTKSAPTTTPKEPAPTTTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 480
QY 439 EPAPTAPKPAPTTTPKEPAPTTTKSAPTTTKESPTTTPKEPAPTTTKSAPTTTKKEP 498
Db 481 EPAPTAPKPAPTTTPKEPAPTTTKSAPTTTKESPTTTPKEPAPTTTKSAPTTTKKEP 540
QY 499 TTKSAPTTTPKEPAPTTTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 558
Db 541 TTKSAPTTTPKEPAPTTTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 600
QY 559 APTAPKPAPTTTPKEPAPTTTKSAPTTTKESPTTTPKEPAPTTTKSAPTTTKKEP 618
Db 601 APTAPKPAPTTTPKEPAPTTTKSAPTTTKESPTTTPKEPAPTTTKSAPTTTKKEP 660
QY 619 PEAPAPTTTPKAAAPNTPEKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 678
Db 661 PEAPAPTTTPKAAAPNTPEKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 720
QY 679 APTTPKPAKELAPTTTKETSTTSKAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTPK 738
Db 721 APTTPKPAKELAPTTTKETSTTSKAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTPK 780
QY 739 TAPTTLKPAPTTTPKPAKELAPTTTKGTSTTSKAPTTTPKGTAPTTTPKEPAPTTPK 798
Db 781 TAPTTLKPAPTTTPKPAKELAPTTTKGTSTTSKAPTTTPKGTAPTTTPKEPAPTTPK 840
QY 799 KPAPTTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAETTPKALENSKPEGVPT 858
Db 841 KPAPTTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAETTPKALENSKPEGVPT 900
QY 859 TKTAAATKPEMTTAKDKTTERDLRTPETTTAAPKMTKETATTTKETSKITATTQV 918
Db 901 TKTAAATKPEMTTAKDKTTERDLRTPETTTAAPKMTKETATTTKETSKITATTQV 960
QY 919 TSTTTQDTPPKITTLTKTLLAPKVTTTKKTTTTEIMNKPETAKPKDRATNSKATTPK 978
Db 961 TSTTTQDTPPKITTLTKTLLAPKVTTTKKTTTTEIMNKPETAKPKDRATNSKATTPK 1020
QY 979 POKTAPKPKTSTKPKTMRVVKPKTTPTRKMTSTMPELAPTSIAEAMLOTTTRN 1038
Db 1021 POKTAPKPKTSTKPKTMRVVKPKTTPTRKMTSTMPELAPTSIAEAMLOTTTRN 1080
QY 1039 QTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMYLPVFNQGIINPMLS 1098
Db 1081 QTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMYLPVFNQGIINPMLS 1140
```

Db 901 TKTPAATKPEMTTAKDKTTERDLRTTTPETTTAAPKMTKETATTTKTESKITATTQV 960
 QY 919 TSTTTQDTPFKITLTKTTLAPKVTTTKKIIITTEIWNKPEETAKPKDRATNSKATTPK 978
 Db 961 TSTTTQDTPFKITLTKTTLAPKVTTTKKIIITTEIWNKPEETAKPKDRATNSKATTPK 1020
 QY 979 POKPTKAPKPTSTKPKTPRVRKPTTTPRKMSTMPBELNPTSRIAEAMLOTTTRPN 1038
 Db 1021 POKPTKAPKPTSTKPKTPRVRKPTTTPRKMSTMPBELNPTSRIAEAMLOTTTRPN 1080
 QY 1039 QTPNSKLVNPKSDAGAGETPHMLLRPHVFPVETTPDMXYLPRVNOGIINPMLS 1098
 Db 1081 QTPNSKLVNPKSDAGAGETPHMLLRPHVFPVETTPDMXYLPRVNOGIINPMLS 1140

RESULT 5
 US-07-757-022B-40
 ; Sequence 40, Application US/07757022B
 ; Patent No. 6433142
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesner, Thomas G.
 ; APPLICANT: Clark, Stephen C.
 ; APPLICANT: Turner, Katherine
 ; APPLICANT: Hewick, Rodney M.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 Cambridgepark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/757,022B
 ; FILING DATE: 19910910
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/546,114
 ; FILING DATE: 29-JUN-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cserr, Luann
 ; REGISTRATION NUMBER: 31,822
 ; REFERENCE/DOCKET NUMBER: GI 5190
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 876-1170
 ; TELEFAX: (617) 876-5851
 ; INFORMATION FOR SEQ ID NO: 40:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1361 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-07-757-022B-40

Query Match 99.7%; Score 5871.9; DB 4; Length 1361;
 Best Local Similarity 99.9%; Pred. No. 5.6e-171;
 Matches 1097; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYGRDATCNCDYNCQHWMECCPDF 60
 Db 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYGRDATCNCDYNCQHWMECCPDF 60
 QY 61 KRVTAEALSCKGRCFSEFERGRCDDAOCKYDKCCPDYSESCAEVHNPTSPSSKAP 120
 Db 61 KRVTAEALSCKGRCFSEFERGRCDDAOCKYDKCCPDYSESCAEVHNPTSPSSKAP 120
 QY 121 PPSGASOTIKSTTKRSPKPNKKTKKVIIESEBITEKVKDNKNKRTKKKPTPKPPVVDEA 180
 Db 121 PPSGASOTIKSTTKRSPKPNKKTKKVIIESEBITEKVKDNKNKRTKKKPTPKPPVVDEA 179
 QY 181 GSGLDNGDFKVTTPDSTTTOHNVKVSPIKTTAKPINRPSLPSPNSDTSKETSLSLVNKET 240
 Db 180 GSGLDNGDFKVTTPDSTTTOHNVKVSPIKTTAKPINRPSLPSPNSDTSKETSLSLVNKET 239
 QY 241 TVETKETTTNNKOTSDGKEKTTSAKETOSIEKTSADKLAPTSKVLAKPTPKAETTTKGP 300
 Db 240 TVETKETTTNNKOTSDGKEKTTSAKETOSIEKTSADKLAPTSKVLAKPTPKAETTTKGP 299
 QY 301 ALTTPEPTTTPKEBPASTTPKPTTTTTSAPTTPKEBPASTTPKAPTTPKBPATTTK 360
 Db 300 ALTTPEPTTTPKEBPASTTPKPTTTTTSAPTTPKEBPASTTPKAPTTPKBPATTTK 359
 QY 361 EPAPTTPKEBPATTTKEBPATTTKSAPTTPKEBPATTPKBPATTPKBPATTPKBPPT 420
 Db 360 EPAPTTPKEBPATTTKEBPATTTKSAPTTPKEBPATTPKBPATTPKBPATTPKBPPT 419
 QY 421 TPKEBPATTTKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBP 480
 Db 420 TPKEBPATTTKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBP 479
 QY 481 APPTTSAPTTPKEBPATTTKSAPTTPKBPPTTPKEBPATTPKBPATTPKBPATTPK 540
 Db 480 APPTTSAPTTPKEBPATTTKSAPTTPKBPPTTPKEBPATTPKBPATTPKBPATTPK 539
 QY 541 EPAPTTPKEBPATTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 600
 Db 540 EPAPTTPKEBPATTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 599
 QY 601 TTPEELAPTTPEBPPTTPEBPATTPKAAAPNTPKBPATTPKBPATTPKBPATTPKBP 660
 Db 600 TTPEELAPTTPEBPPTTPEBPATTPKAAAPNTPKBPATTPKBPATTPKBPATTPKBP 659
 QY 661 ETAPTTKGTAPTTLKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 720
 Db 660 ETAPTTKGTAPTTLKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 719
 QY 721 KEPAATTPKBPATTPKGTAPTTLKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 780
 Db 720 KEPAATTPKBPATTPKGTAPTTLKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 779
 QY 781 PKETAPTTKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 840
 Db 780 PKETAPTTKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 839
 QY 841 EPTPKALENSPKBPVPTTKTTPAATKPEMTTTAKDTERDLRTTPTTTTAAAPKMTKETA 900
 Db 840 EPTPKALENSPKBPVPTTKTTPAATKPEMTTTAKDTERDLRTTPTTTTAAAPKMTKETA 899
 QY 901 TTTTEKTESKITATTQVTSITTTQDTPFKITLTKTTLAPKVTTTKKIIITTEIWNKPE 960
 Db 900 TTTTEKTESKITATTQVTSITTTQDTPFKITLTKTTLAPKVTTTKKIIITTEIWNKPE 959
 QY 961 ETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTPRVRKPTTTPRKMSTMPBEL 1020
 Db 960 ETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTPRVRKPTTTPRKMSTMPBEL 1019
 QY 1021 NPTSRIAEAMLOTTTRNPTNSKLVNPKSDAGAGETPHMLLRPHVFPVETTPDM 1080
 Db 1020 NPTSRIAEAMLOTTTRNPTNSKLVNPKSDAGAGETPHMLLRPHVFPVETTPDM 1079

```

QY 1081 DYLPVNOGIIINPMLS 1098
Db 1080 DYLPVNOGIIINPMLS 1097

RESULT 6
US-10-164-595-78
; Sequence 78, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-78

Query Match 99.6%; Score 5863.8; DB 4; Length 1404;
Best Local Similarity 96.1%; Pred. No. 1e-170;
Matches 1096; Conservative 0; Mismatches 2; Indels 42; Gaps 1;

QY 1 MAWKLPYLLLLSVFVIQQVSSODLSSCAGRCGEGYSRDATCNCYNOCHYMECCPDF 60
Db 1 MAWKLPYLLLLSVFVIQQVSSODLSSCAGRCGEGYSRDATCNCYNOCHYMECCPDF 60

QY 61 KRVTAEILSCGRCPESPERGECDCDAQCKYDKCCPDYDFCAEAVHNPTSPSSKKAP 120
Db 61 KRVTAEILSCGRCPESPERGECDCDAQCKYDKCCPDYDFCAEAVHNPTSPSSKKAP 120

QY 121 PPSGASQIKSTTKRSPKPPNKKTKKVIIEEITE----- 156
Db 121 PPSGASQIKSTTKRSPKPPNKKTKKVIIEEITEHSVSENQSSSSSSSSSSSTIR 180

QY 157 -----KVQDKNKNRTKKPTKPPVVDGAGSLGDFKVTPTST 198
Db 181 KIKSKNAAREQOKLKVQDKNKNRTKKPTKPPVVDGAGSLGDFKVTPTST 240

QY 199 TQHNKVSPTKITTAKPINPRLPPNSDTSKETSLSLVNKETTVETKETTNNKQSTDG 258
Db 241 TQHNKVSPTKITTAKPINPRLPPNSDTSKETSLSLVNKETTVETKETTNNKQSTDG 300

QY 259 KEKTTSAKETOSIEKTSKADLAPTSKULAKPTPKAETTKGPAITTKGPAITTKGPA 318
Db 301 KEKTTSAKETOSIEKTSKADLAPTSKVLAKPTPKAETTKGPAITTKGPAITTKGPA 360

QY 319 TTPKEPTTIKSAPTTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPA 378
Db 361 TTPKEPTTIKSAPTTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPA 420

QY 379 APTTKSAPTTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPA 438
Db 421 APTTKSAPTTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPA 480

QY 439 EPAPTAPKPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPA 498
Db 481 EPAPTAPKPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPA 540

QY 499 TTKSAPTTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPA 558
Db 541 TTKSAPTTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPA 600

QY 559 APTAPKPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPA 618
Db 601 APTAPKPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPA 660

QY 619 PEEPAITTKAAAPNTPKEPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPAITTKGPA 678

```

RESULT 7

```

US-07-757-022B-48
; Sequence 48, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822

```

```

; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1354 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-48

Query Match          95.9%; Score 5647.2; DB 4; Length 1354;
Best Local Similarity 96.3%; Pred. No. 3.8e-164;
Matches 1057; Conservative 7; Mismatches 26; Indels 8; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60

QY 61 KRVCTAEISCKGRCFESFERGECDCDAQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120
DB 61 KRVCTAEISCKGRCFESFERGECDCDAQCKYDKCCPDYSEFCAEHSVGENQESSSS 120

QY 121 PPSGASQITKSTTKSPKPPNKKTKVIBSEITEKVKDNKNKNTKKKPTKPPVDEA 180
DB 121 SSSSSSSSTIWKISSKNSAARELOKLL-----KVKDNKNKNTKKKPTKPPVDEA 172

QY 181 GSGLDNGPQVTTPTDSTQHNKVSPTKITTAKPINRPSLPNSDTSKETSIVNKET 240
DB 173 GSGLDNGDPKVITPTDSTQHNKVSPTKITTAKPINRPSLPNSDTSKETSIVNKET 232

QY 241 TVETKETTITNKQTDGKEKTSKAKTQSIKTSKADLAPTSKVLAKPTPKAETTTKGP 300
DB 233 TVETKETTITNKQTDGKEKTSKAKTQSIKTSKADLAPTSKVLAKPTPKAETTTKGP 292

QY 301 ALTTPEKPTPTTKPEASTTKPEPTTKISAPTTKPEAPTTKSAPTTKPEAPTTTK 360
DB 293 ALTTPEKPTPTTKPEASTTKPEPTTKISAPTTKPEAPTTKSAPTTKPEAPTTTK 352

QY 361 EPAPTTPEKAPTTTKPEAPTTTKSAPTTKPEAPTTPKKAPTTTKPEAPTTKPEPT 420
DB 353 EPAPTTPEKAPTTTKPEAPTTTKSAPTTKPEAPTTPKKAPTTTKPEAPTTKPEPT 412

QY 421 TPKEAPTTKEAPTTTKPEAPTTAPKAPPTTKPEAPTTKPEAPTTTKPEPTTKPEP 480
DB 413 TPKEAPTTKEAPTTTKPEAPTTAPKAPPTTKPEAPTTKPEAPTTTKPEPTTKPEP 472

QY 481 APPTTKSAPTTKEAPTTTKSAPTTKPEPSPTTKPEAPTTKPEAPTTPKKAPTTPK 540
DB 473 APPTTKSAPTTKEAPTTTKSAPTTKPEPSPTTKPEAPTTKPEAPTTPKKAPTTPK 532

QY 541 EPAPTTPEKAPTTTKKAPTAPEAPTTKPEAPTTPKETAPTTPKLTPTTPEKLAPTTPEKAP 600
DB 533 EPAPTTPEKAPTTTKKAPTAPEAPTTKPEAPTTPKETAPTTPKLTPTTPEKLAPTTPEKAP 592

QY 601 TPEELAPPTPEPTTTEEBAPTTKAAANTKPEAPTTKPEAPTTKPEAPTTKPEAPTTK 660
DB 593 TPEELAPPTPEPTTTEEBAPTTKAAANTKPEAPTTKPEAPTTKPEAPTTKPEAPTTK 652

QY 661 ETAPTTPKGTAPTTLKPEAPTTPKDAPKELAPTTTKPTSTTSKAPTTKGTAPTTT 720
DB 653 ETAPTTPKGTAPTTLKPEAPTTPKDAPKELAPTTTKPTSTTSKAPTTKGTAPTTT 712

QY 721 KEAPTTPEKAPTTTKGTAPTTLKPEAPTTPKDAPKELAPTTTKGPTSTTSKAPTT 780
DB 713 KEAPTTPEKAPTTTKGTAPTTLKPEAPTTPKDAPKELAPTTTKGPTSTTSKAPTT 772

QY 781 PKETAPTTKEAPTTPKAPPTTETPPPTTSEVSTPTTKETPTTIHKSPESTPELSA 840
DB 773 PKETAPTTKEAPTTPKAPPTTETPPPTTSEVSTPTTKETPTTIHKSPESTPELSA 832

QY 841 EPTPKALENSPKPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPTTAAAPKMTKETA 900

```

```

RESULT 8
US-07-757-022B-52
; Sequence 52, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1363 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-52

```

Query Match 95.3%; Score 5608.7; DB 4; Length 1363;
Best Local Similarity 92.7%; Pred. No. 5.6e-163;
Matches 1057; Conservative 0; Mismatches 0; Indels 83; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSDATCNCDCYNCHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25

QY 61 KRVCTAELSCRCRCPESFERGECDCDAQCKYDKCCDDYSPCAEVENPTSPSSKAP 120
DB -----ELSCRCRCPESFERGECDCDAQCKYDKCCDDYSPCAEVENPTSPSSKAP 79

QY 121 PPSGASQITKSTTKSPKPPNKKTKKVIKIESEITE----- 156
DB PPSGASQITKSTTKSPKPPNKKTKKVIKIESEITEHSVSENQESSSSSSSSSSSTI 139

QY 157 -----KVONKKNRTKKKPTPKPPVVDAGSLDNGDFKVTTPDTST 198
DB KIKSSKNSAANRELQKLVKONKKNRTKKKPTPKPPVVDAGSLDNGDFKVTTPDTST 199

QY 199 TQHNKVSPTKITTAKPINRSLPNSDTSKETSITVKNKSTTVETKETTNNKQTSDDG 258
DB TQHNKVSPTKITTAKPINRSLPNSDTSKETSITVKNKSTTVETKETTNNKQTSDDG 259

QY 259 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKETPTTTPKEPAS 318
DB KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKETPTTTPKEPAS 319

QY 319 TTPKEPTTTIKSAPTTKEPAPTTTKGAPTTKEPAPTTTKGAPTTKEPAPTTTKGAPTT 378
DB TTPKEPTTTIKSAPTTKEPAPTTTKGAPTTKEPAPTTTKGAPTTKEPAPTTTKGAPTT 379

QY 379 APTTTKSAPTTKEPAPTTTKGAPTTKEPAPTTTKGAPTTKEPAPTTTKGAPTTTKP 438
DB APTTTKSAPTTKEPAPTTTKGAPTTKEPAPTTTKGAPTTKEPAPTTTKGAPTTTKP 439

QY 439 EPAPTAPKAPPTTKGAPTTKEPAPTTTKGAPTTKEPAPTTTKGAPTTTKGAPTTKEPAPT 498
DB EPAPTAPKAPPTTKGAPTTKEPAPTTTKGAPTTKEPAPTTTKGAPTTTKGAPTTKEPAPT 499

QY 499 TTKSAPTTKEPAPTTTKGAPTTKEPAPTTTKGAPTTKEPAPTTTKGAPTTTKGAPTT 558
DB TTKSAPTTKEPAPTTTKGAPTTKEPAPTTTKGAPTTKEPAPTTTKGAPTTTKGAPTT 559

QY 559 APTAPKEPAPTTKEPAPTTTKGAPTTKEPAPTTTKGAPTTKEPAPTTKEPAPTTKEPAPT 618
DB APTAPKEPAPTTKEPAPTTTKGAPTTKEPAPTTTKGAPTTKEPAPTTKEPAPTTKEPAPT 619

QY 619 PEEPAPTTKAAAPNTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEP 678
DB PEEPAPTTKAAAPNTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEP 679

QY 679 APTTPKXPAKELAPTTTKETSTSDKAPTTTPKGATPTTPKEPAPTTKEPAPTTPKG 738
DB APTTPKXPAKELAPTTTKETSTSDKAPTTTPKGATPTTPKEPAPTTKEPAPTTPKG 739

QY 739 TAPTTLKEPAPTTKAPKAPKELAPTTTKGPTSTSDKAPTTTPKETAPTTPKGTAPTTLKEP 798
DB TAPTTLKEPAPTTKAPKAPKELAPTTTKGPTSTSDKAPTTTPKETAPTTPKGTAPTTLKEP 799

QY 799 KPAPTTTPPTTSEVSTPTTTKEPTTIHKSPPDESTPELSAETPKALENSKPEGVP 858
DB KPAPTTTPPTTSEVSTPTTTKEPTTIHKSPPDESTPELSAETPKALENSKPEGVP 859

QY 859 TKTPAATKPEMTTAKDKTTERDLRTPPETTAPAKMTKETATTEKTESKITATTQV 918
DB TKTPAATKPEMTTAKDKTTERDLRTPPETTAPAKMTKETATTEKTESKITATTQV 919

QY 919 TSTTTQDTTPPKITTLKTTTLAPKVTITTKTITTEIMNKPEETAKPKDRATNSKATTPK 978
DB TSTTTQDTTPPKITTLKTTTLAPKVTITTKTITTEIMNKPEETAKPKDRATNSKATTPK 979

QY 979 POKPTKAPKPTSTKPKTMPRVKPKTTPTPKOTSTMPELNPTSRIAEAMLOTTTREN 1038
DB 980 POKPTKAPKPTSTKPKTMPRVKPKTTPTPKOTSTMPELNPTSRIAEAMLOTTTREN 1039

QY 1039 QTNENKLVNPKSDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1098
DB 1040 QTNENKLVNPKSDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1099

RESULT 9
US-07-757-022B-42
; Sequence 42, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Geener, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserk, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1311 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-42

Query Match 95.3%; Score 5607.9; DB 4; Length 1311;
Best Local Similarity 95.4%; Pred. No. 5.7e-163;
Matches 1047; Conservative 0; Mismatches 0; Indels 51; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSDATCNCDCYNCHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSDATCNCDCYNCHYMECCPDF 60

QY 61 KRVCTAELSCRCRCPESFERGECDCDAQCKYDKCCDDYSPCAEVENPTSPSSKAP 120

Db 61 KRVTAEISCKGRCPESFRGREGDCDAQCKYDKCCPDYBSFCAE----- 106
QY 121 PPASQTIKSTTKSPKPNKKTKKVIKIESEBETKVKDNKNKNTKKKTPKPPVDEA 180
Db 107 -----VKDNKNKNTKKKTPKPPVDEA 129
QY 181 GSGLDNGDFKVTTPDSTTQHNKVSSTPKITTAKEPINRPSLPPNSDTSKETSLSLVNKET 240
Db 130 GSGLDNGDFKVTTPDSTTQHNKVSSTPKITTAKEPINRPSLPPNSDTSKETSLSLVNKET 189
QY 241 TVETKETTITNKQSTGDKKEKTSKAKETQSIKTSKAKOLAPTSKVLAKPTPKAETTTKGP 300
Db 190 TVETKETTITNKQSTGDKKEKTSKAKETQSIKTSKAKOLAPTSKVLAKPTPKAETTTKGP 249
QY 301 ALTTKEPTPTTKPEPATTPKEPTPTTKSAPTTKPEPATTTKSAATTKPEPATTTK 360
Db 250 ALTTKEPTPTTKPEPATTPKEPTPTTKSAPTTKPEPATTTKSAATTKPEPATTTK 309
QY 361 EPAPTTKPEPATTTKPEPATTTKGAFTTKPEPATTPKKPAPTTPKEPATTTKPEPT 420
Db 310 EPAPTTKPEPATTTKPEPATTTKGAFTTKPEPATTPKKPAPTTPKEPATTTKPEPT 369
QY 421 TPKEPATTTKPEPATTPKEPATTPKAPATTPKAPATTPKAPATTPKAPATTPK 480
Db 370 TPKEPATTTKPEPATTPKAPATTPKAPATTPKAPATTPKAPATTPKAPATTPK 429
QY 481 APPTTKSAPTTKPEPATTTKAPATTPKAPATTPKAPATTPKAPATTPKAPATTPK 540
Db 430 APPTTKSAPTTKPEPATTTKAPATTPKAPATTPKAPATTPKAPATTPKAPATTPK 489
QY 541 EPAPTTKPEPATTTTKAPATTPKAPATTPKAPATTPKAPATTPKAPATTPKAP 600
Db 490 EPAPTTKPEPATTTTKAPATTPKAPATTPKAPATTPKAPATTPKAPATTPKAP 549
QY 601 TPPELAPTTPEEPPTTPEEPAPTTPKAAANTKPEPATTPKAPATTPKAPATTPK 660
Db 550 TPPELAPTTPEEPPTTPEEPAPTTPKAAANTKPEPATTPKAPATTPKAPATTPK 609
QY 661 ETAPTTPKGTATTLKEPATTPKAPKELAPTTTKPTSTTSKAPATTPKGTAPTT 720
Db 610 ETAPTTPKGTATTLKEPATTPKAPKELAPTTTKPTSTTSKAPATTPKGTAPTT 669
QY 721 KEPATTPKAPATTPKGTAPTTKELAPATTPKAPKELAPATTTKGTSTTSKAPATTT 780
Db 670 KEPATTPKAPATTPKGTAPTTKELAPATTPKAPKELAPATTTKGTSTTSKAPATTT 729
QY 781 PKETAPTTKPEPATTPKAPATTPETPPPTTSEVSTPTTKPTTTHKSPDESPELSA 840
Db 730 PKETAPTTKPEPATTPKAPATTPETPPPTTSEVSTPTTKPTTTHKSPDESPELSA 789
QY 841 EPTPKALENSPKPGVPTTKTPAATKPEMTTAKDKTTERDLRTTPTTTAAPKMTKETA 900
Db 790 EPTPKALENSPKPGVPTTKTPAATKPEMTTAKDKTTERDLRTTPTTTAAPKMTKETA 849
QY 901 TTTKTTESKITTATQVSTTQDTTPPKITTLTKTTTILAPKVTTKITTTTILMNKPE 960
Db 850 TTTKTTESKITTATQVSTTQDTTPPKITTLTKTTTILAPKVTTKITTTTILMNKPE 909
QY 961 ETAPKDRATNSKATTPKPKETKAPKPTSTTKPKPTWPRVAKPTTTPTRKMTSTMPEL 1020
Db 910 ETAPKDRATNSKATTPKPKETKAPKPTSTTKPKPTWPRVAKPTTTPTRKMTSTMPEL 969
QY 1021 NPTSRIAEAMLOTTTRPNQTPNSKLVNPKSEDAGGAEGETPHMLLRPHVMPBVTPTDM 1080
Db 970 NPTSRIAEAMLOTTTRPNQTPNSKLVNPKSEDAGGAEGETPHMLLRPHVMPBVTPTDM 1029
QY 1081 DYLPRVNOGIIINPMLS 1098
Db 1030 DYLPRVNOGIIINPMLS 1047

RESULT 10

US-07-757-022B-46

/ Sequence 46, Application US/07757022B
/ Patent No. 6433142
/ GENERAL INFORMATION:
/ APPLICANT: Gesner, Thomas G.
/ APPLICANT: Clark, Stephen C.
/ APPLICANT: Turner, Katherine
/ APPLICANT: Hewick, Rodney M.
/ TITLE OF INVENTION: Megakaryocyte Stimulating Factors
/ NUMBER OF SEQUENCES: 143
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Genetics Institute, Inc.
/ STREET: 87 CambridgePark Drive
/ CITY: Cambridge
/ STATE: Massachusetts
/ COUNTRY: U.S.A.
/ ZIP: 02140
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/07/757,022B
/ FILING DATE: 19910910
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/643,502
/ FILING DATE: 18-JAN-1991
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/546,114
/ FILING DATE: 29-JUN-1990
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/457,196
/ FILING DATE: 29-DEC-1989
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/390,901
/ FILING DATE: 08-AUG-1989
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Cserr, Luann
/ REGISTRATION NUMBER: 31,822
/ REFERENCE/DOCKET NUMBER: GI 5190
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617)876-1170
/ TELEFAX: (617)876-5851
/ INFORMATION FOR SEQ ID NO: 46:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1320 amino acids
/ TYPE: AMINO ACID
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-07-757-022B-46

Query Match 95.3%; Score 5607.8; DB 4; Length 1320;
Best Local Similarity 96.2%; Pred. No. 5.8e-163;
Matches 1056; Conservative 0; Mismatches 0; Indels 42; Gaps 2;

QY 1 MANKTLPIYLLLLSVFVIQQVSSQDLSSCAGREGYSDATCNCDYNQHYNECCPDF 60
Db 1 MANKTLPIYLLLLSVFVIQQVSSQ----- 25
QY 61 KRVTAEISCKGRCPESFRGREGDCDAQCKYDKCCPDYBSFCAEVHNPTSPSSKKAP 120
Db 26 -----ELSCKGRCPESFRGREGDCDAQCKYDKCCPDYBSFCAEVHNPTSPSSKKAP 79
QY 121 PPASQTIKSTTKSPKPNKKTKKVIKIESEBETKVKDNKNKNTKKKTPKPPVDEA 180
Db 80 PPASQTIKSTTKSPKPNKKTKKVIKIESEBETKVKDNKNKNTKKKTPKPPVDEA 138
QY 181 GSGLDNGDFKVTTPDSTTQHNKVSSTPKITTAKEPINRPSLPPNSDTSKETSLSLVNKET 240
Db 139 GSGLDNGDFKVTTPDSTTQHNKVSSTPKITTAKEPINRPSLPPNSDTSKETSLSLVNKET 198
QY 241 TVETKETTITNKQSTGDKKEKTSKAKETQSIKTSKAKOLAPTSKVLAKPTPKAETTTKGP 300

```

199 TVETKETTNNKQSTDCGKNTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGP 258
301 ALTTKKEPTTTPKEPASSTPKKEPTTTIKSAPTTKEPAPTTTKSAPTTKEPAPTTTK 360
259 ALTTKKEPTTTPKEPASSTPKKEPTTTIKSAPTTKEPAPTTTKSAPTTKEPAPTTTK 318
361 EPAPTTKEPAPTTTKKAPTTTKSAPTTKEPAPTTTKKAPTTTKKAPTTTKKAPTTTK 420
319 EPAPTTKEPAPTTTKKAPTTTKSAPTTKEPAPTTTKKAPTTTKKAPTTTKKAPTTTK 378
421 TKKEPAPTTKGPAPTTKEPAPTTAPKAPTTAPKAPTTAPKAPTTAPKAPTTAPKAPTT 480
379 TKKEPAPTTKGPAPTTKEPAPTTAPKAPTTAPKAPTTAPKAPTTAPKAPTTAPKAPTT 438
481 APPTTKSAPTTTKKAPTTTKSAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTT 540
439 APPTTKSAPTTTKKAPTTTKSAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTT 498
541 EPAPTTKEPAPTTTKKAPTTAPKAPTTAPKAPTTAPKAPTTAPKAPTTAPKAPTTAPKAPTT 600
499 EPAPTTKEPAPTTTKKAPTTAPKAPTTAPKAPTTAPKAPTTAPKAPTTAPKAPTTAPKAPTT 558
601 TTPESLAPTTKEEPTTTPKEPAPTTTPKAAAPNTPKAPTTTPKAPTTTPKAPTTTPKAPTT 660
559 TTPESLAPTTKEEPTTTPKEPAPTTTPKAAAPNTPKAPTTTPKAPTTTPKAPTTTPKAPTT 618
661 ETAPTTTPKGTAPTTLKEPAPTTTPKAPKAPKELAPTTTKEPTTSTSDKAPTTTPKGTATTP 720
619 ETAPTTTPKGTAPTTLKEPAPTTTPKAPKAPKELAPTTTKEPTTSTSDKAPTTTPKGTATTP 678
721 KEAPTTTPKAPTTTPKGTATTPKAPKAPKELAPTTTKEPTTSTSDKAPTTTPKGTATTP 780
679 KEAPTTTPKAPTTTPKGTATTPKAPKAPKELAPTTTKEPTTSTSDKAPTTTPKGTATTP 738
781 PKETAPTTTPKAPTTTPKAPKAPKELAPTTTKEPTTSTSDKAPTTTPKGTATTP 840
739 PKETAPTTTPKAPTTTPKAPKAPKELAPTTTKEPTTSTSDKAPTTTPKGTATTP 798
841 EPTPKALENSKPEGPVPTTKTAPKAPKAPKELAPTTTKEPTTSTSDKAPTTTPKGTATTP 900
799 EPTPKALENSKPEGPVPTTKTAPKAPKAPKELAPTTTKEPTTSTSDKAPTTTPKGTATTP 858
901 TTTKTTESKLTATTTQVSTTQDTPKLTTLKTTTLAPKVTITTKTITTTIMNKPE 960
859 TTTKTTESKLTATTTQVSTTQDTPKLTTLKTTTLAPKVTITTKTITTTIMNKPE 918
961 ETAPKPKDRATNSKATTPKPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 1020
919 ETAPKPKDRATNSKATTPKPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 978
1021 NPTSIARAMLQTTTRPNQTPNSKLVEVNPSEBAGGEGTTPHMLRPHVFMPEVTPDM 1080
979 NPTSIARAMLQTTTRPNQTPNSKLVEVNPSEBAGGEGTTPHMLRPHVFMPEVTPDM 1038
1081 DYLPRVNPQGGIINPMLS 1098
1039 DYLPRVNPQGGIINPMLS 1056

```

RESULT 11

```

US-07-757-022B-60
; Sequence 60, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.

```

```

; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-60

```

```

Query Match 95.3%; Score 5607.8; DB 4; Length 1320;
Best Local Similarity 96.2%; Pred. No. 5.8e-163;
Matches 1056; Conservative 0; Mismatches 0; Indels 42; Gaps 2;

```

```

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSCAGRCGEGYSDATCNDYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
QY 61 KRVCYTAELSCGRCSPESFERGECDCDAQCKKYDKCCPDYBSFCAEVHNP'TSPSSKKAP 120
DB 26 -----ELSCGRCSPESFERGECDCDAQCKKYDKCCPDYBSFCAEVHNP'TSPSSKKAP 79
QY 121 PPSGASQTIKSTTKSPKPPNKKTKKVIKESBEIEKVKDKNKNKTKKTKKTPKPPVDEA 180
DB 80 PPSGASQTIKSTTKSPKPPNKKTKKVIKESBEIEKVKDKNKNKTKKTKKTPKPPVDEA 138
QY 181 GSGLDNGDFKVTTPDTSTTQHNKSVTSKPIITAKINPRPSLPNSDTSKETSITVNET 240
DB 139 GSGLDNGDFKVTTPDTSTTQHNKSVTSKPIITAKINPRPSLPNSDTSKETSITVNET 198
QY 241 TVETKTTTTNKTSTGDKKETSIAKTSQTSIEKTSKADLAPTSKVLAKPTPKAETTTKGP 300
DB 199 TVETKTTTTNKTSTGDKKETSIAKTSQTSIEKTSKADLAPTSKVLAKPTPKAETTTKGP 258
QY 301 ALTTKKEPTTTPKBPASTTPKEPTTTIKSAPTTKEPAPTTTKSAPTTKEPAPTTTK 360
DB 259 ALTTKKEPTTTPKBPASTTPKEPTTTIKSAPTTKEPAPTTTKSAPTTKEPAPTTTK 318
QY 361 EPAPTTKEPAPTTTKKAPTTTKSAPTTKEPAPTTTKKAPTTTKKAPTTTKKAPTTTK 420
DB 319 EPAPTTKEPAPTTTKKAPTTTKSAPTTKEPAPTTTKKAPTTTKKAPTTTKKAPTTTK 378

```


RESULT 13
 US-07-757-022B-84
 ; Sequence 84, Application US/07757022B
 ; Patent No. 6433142
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesner, Thomas G.
 ; APPLICANT: Clark, Stephen C.
 ; APPLICANT: Turner, Katherine
 ; APPLICANT: Hewick, Rodney M.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/757,022B
 ; FILING DATE: 19910910
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/546,114
 ; FILING DATE: 29-JUN-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Csert, Luann
 ; REGISTRATION NUMBER: 31,822
 ; REFERENCE/DOCKET NUMBER: GI 5190
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 876-1170
 ; TELEFAX: (617) 876-5851
 ; INFORMATION FOR SEQ ID NO: 84:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1022 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-07-757-022B-84

Query Match 92.9%; Score 5466.9; DB 4; Length 1022;
 Best Local Similarity 95.1%; Pred. No. 8e-159;
 Matches 1020; Conservative 0; Mismatches 2; Indels 51; Gaps 2;
 QY 26 DLSSCAGCGGGRDATFNCNDYNCQHYMECCPDKRVCTAELSCGRCFESFERGRECD 85
 DB 1 DLSSCAGCGGGRDATFNCNDYNCQHYMECCPDKRVCTAELSCGRCFESFERGRECD 58
 QY 86 CDACKKYDKCCPYESCAEVHNPFPSPSSKXAPPPSGASQITKSTTKSPKPPNKKY 145
 DB 59 CDACKKYDKCCPYESCAEVHNPFPSPSSKXAPPPSGASQITKSTTKSPKPPNKKY 80
 QY 146 KKVIESEITEKVKDKNKKRTKKKTPKPPVVDGAGSLDNGDFKVTTPDTSTTQHNKVS 205
 DB 81 -----AVKDNKKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDTSTTQHNKVS 129
 QY 206 TSPKLTAKPINRPSLPNSDTSKTSLVNKETTVEKTTTNNKQSTDGKEKTTSA 265

DB 130 TSPKLTAKPINRPSLPNSDTSKTSLVNKETTVEKTTTNNKQSTDGKEKTTSA 189
 QY 266 KETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPASTTTPKEPT 325
 DB 190 KETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPASTTTPKEPT 249
 QY 326 PTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKS 385
 DB 250 PTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKS 309
 QY 386 APTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 445
 DB 310 APTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 369
 QY 446 KKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 505
 DB 370 KKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 429
 QY 506 TPKEPSPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 565
 DB 430 TPKEPSPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 489
 QY 566 PAPTTPKETAPTTPKLTPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 625
 DB 490 PAPTTPKETAPTTPKLTPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 549
 QY 626 TPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 685
 DB 550 TPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 609
 QY 686 PAPKELAPTTTPKEPTSTTSKDPAPTTPKGTAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 745
 DB 610 PAPKELAPTTTPKEPTSTTSKDPAPTTPKGTAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 669
 QY 746 EPAPTTPKPAKELAPTTTPKGTSTTSKDPAPTTPKGTAPTTPKEPAPTTTPKEPAPTTTPKEPAP 805
 DB 670 EPAPTTPKPAKELAPTTTPKGTSTTSKDPAPTTPKGTAPTTPKEPAPTTTPKEPAPTTTPKEPAP 729
 QY 806 ETPTPTTSEVSTPTTPKETTTHKSPDESTPSELSAETPKALENSPKGPGVPTTKTTPAAT 865
 DB 730 ETPTPTTSEVSTPTTPKETTTHKSPDESTPSELSAETPKALENSPKGPGVPTTKTTPAAT 789
 QY 866 KPEMTTAKDKTTERDLRTPPTTAAAPKVTKEATTTTEKTTESKITATTQVTTSTTQD 925
 DB 790 KPEMTTAKDKTTERDLRTPPTTAAAPKVTKEATTTTEKTTESKITATTQVTTSTTQD 849
 QY 926 TTPFKITLTKTTLAPKVTTKKTTTTEIMNKPEETAKPKDRATNSKATTPKPKPTKA 985
 DB 850 TTPFKITLTKTTLAPKVTTKKTTTTEIMNKPEETAKPKDRATNSKATTPKPKPTKA 909
 QY 986 PKKPTSTKKPKTMVRKPKTTTPPKMTSTNPELNPTSRIAEAMLTTPRNPQTPNSKL 1045
 DB 910 PKKPTSTKKPKTMVRKPKTTTPPKMTSTNPELNPTSRIAEAMLTTPRNPQTPNSKL 969
 QY 1046 VEVPKSEDAGGAGETPHMLRPHVFMPEVTPDMDYLPVFNQGIILNPMLS 1098
 DB 970 VEVPKSEDAGGAGETPHMLRPHVFMPEVTPDMDYLPVFNQGIILNPMLS 1022

RESULT 14
 US-07-757-022B-50
 ; Sequence 50, Application US/07757022B
 ; Patent No. 6433142
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesner, Thomas G.
 ; APPLICANT: Clark, Stephen C.
 ; APPLICANT: Turner, Katherine
 ; APPLICANT: Hewick, Rodney M.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.

```

; STREET: 87 Cambridgepark Drive
; City: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1314 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-50

Query Match          91.7%; Score 5399.2; DB 4; Length 1314;
Best Local Similarity 92.7%; Pred. No. 1.3e-156;
Matches 1018; Conservative 7; Mismatches 25; Indels 48; Gaps 2;

QY      1  MAWKTLPIYLLLLSVFVYQVSSQDLSSCAGRCGEGYSRDATCNDYNCQHYMECCPDF 60
DB      1  MAWKTLPIYLLLLSVFVYQVSSQDLSSCAGRCGEGYSRDATCNDYNCQHYMECCPDF 60

QY      61  KRVTAEIISCKGRCFESFERGRECDCAQCKYDKCCPDYSEFCAEVHNPTSPSPSSKKAP 120
DB      61  KRVTAEI-----HSVSENQSSSSS 80

QY      121  PFGASQTKSTKSPKPPNKKTKKVIIESEITEKVDNKNKTKKKPTPKPPVUDEA 180
DB      121  PFGASQTKSTKSPKPPNKKTKKVIIESEITEKVDNKNKTKKKPTPKPPVUDEA 180

QY      81  SSSSSSTWKIKSSKNSAANLEQKL-----KVDNKNKTKKKPTPKPPVUDEA 132
DB      81  SSSSSSTWKIKSSKNSAANLEQKL-----KVDNKNKTKKKPTPKPPVUDEA 132

QY      181  GSGLDNGDKVTPDSTTQHNKVSPTSITAKPINRPSLPNSDTSKETSILVNKET 240
DB      181  GSGLDNGDKVTPDSTTQHNKVSPTSITAKPINRPSLPNSDTSKETSILVNKET 240

QY      193  GSGLDNGDKVTPDSTTQHNKVSPTSITAKPINRPSLPNSDTSKETSILVNKET 192
DB      193  GSGLDNGDKVTPDSTTQHNKVSPTSITAKPINRPSLPNSDTSKETSILVNKET 192

QY      241  TVETKTTTTNKQTSIDGKEKTSKETSIEKTSIAKDLAPTSKVLAKPTPKAETTKGP 300
DB      241  TVETKTTTTNKQTSIDGKEKTSKETSIEKTSIAKDLAPTSKVLAKPTPKAETTKGP 300

QY      193  TVETKTTTTNKQTSIDGKEKTSKETSIEKTSIAKDLAPTSKVLAKPTPKAETTKGP 252
DB      193  TVETKTTTTNKQTSIDGKEKTSKETSIEKTSIAKDLAPTSKVLAKPTPKAETTKGP 252

QY      301  ALTTTPKEPTTPKPEASTTKEPTTTIKSAPTTPKEAPATTTKSAPTTPKEAPATTK 360
DB      301  ALTTTPKEPTTPKPEASTTKEPTTTIKSAPTTPKEAPATTTKSAPTTPKEAPATTK 360

QY      253  ALTTTPKEPTTPKPEASTTKEPTTTIKSAPTTPKEAPATTTKSAPTTPKEAPATTK 312
DB      253  ALTTTPKEPTTPKPEASTTKEPTTTIKSAPTTPKEAPATTTKSAPTTPKEAPATTK 312

QY      361  EPAPTTTPKEAPATTTKSAPTTPKKEPAPTTPKKPAFTTPKKEPAPTTKKEPTPT 420
DB      361  EPAPTTTPKEAPATTTKSAPTTPKKEPAPTTPKKPAFTTPKKEPAPTTKKEPTPT 420

QY      313  EPAPTTTPKEAPATTTKSAPTTPKKEPAPTTPKKPAFTTPKKEPAPTTKKEPTPT 372
DB      313  EPAPTTTPKEAPATTTKSAPTTPKKEPAPTTPKKPAFTTPKKEPAPTTKKEPTPT 372

QY      421  TPKEPAPTTTPKEPAPTTAPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 480
DB      421  TPKEPAPTTTPKEPAPTTAPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 480

QY      373  TPKEPAPTTTPKEPAPTTAPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 432
DB      373  TPKEPAPTTTPKEPAPTTAPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 432

QY      481  APITTKSAPTTTPKEPAPTTTPKSAPTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTPK 540
DB      481  APITTKSAPTTTPKEPAPTTTPKSAPTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTPK 540

QY      433  APITTKSAPTTTPKEPAPTTTPKSAPTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTPK 492
DB      433  APITTKSAPTTTPKEPAPTTTPKSAPTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTPK 492

QY      541  EPAPTTTPKEPAPTTTPKKEPAPTTTPKSAPTTPKKEPAPTTTPKKEPAPTTTPKKEPAP 600
DB      541  EPAPTTTPKEPAPTTTPKKEPAPTTTPKSAPTTPKKEPAPTTTPKKEPAPTTTPKKEPAP 600

QY      493  EPAPTTTPKEPAPTTTPKKEPAPTTTPKSAPTTPKKEPAPTTTPKKEPAPTTTPKKEPAP 552
DB      493  EPAPTTTPKEPAPTTTPKKEPAPTTTPKSAPTTPKKEPAPTTTPKKEPAPTTTPKKEPAP 552

QY      601  TPPELAPTTPEEPTTPPEEAPATTPKAAAPNTPKEPAPTTTPKKEPAPTTTPKKEPAPTTPK 660
DB      601  TPPELAPTTPEEPTTPPEEAPATTPKAAAPNTPKEPAPTTTPKKEPAPTTTPKKEPAPTTPK 660

QY      553  TPPELAPTTPEEPTTPPEEAPATTPKAAAPNTPKEPAPTTTPKKEPAPTTTPKKEPAPTTPK 612
DB      553  TPPELAPTTPEEPTTPPEEAPATTPKAAAPNTPKEPAPTTTPKKEPAPTTTPKKEPAPTTPK 612

QY      661  ETAPTTTPKGTAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKGTATTP 720
DB      661  ETAPTTTPKGTAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKGTATTP 720

QY      613  ETAPTTTPKGTAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKGTATTP 672
DB      613  ETAPTTTPKGTAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKGTATTP 672

QY      721  KEPAPTTTPKGTAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKGTATTP 780
DB      721  KEPAPTTTPKGTAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKGTATTP 780

QY      673  KEPAPTTTPKGTAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKGTATTP 732
DB      673  KEPAPTTTPKGTAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKGTATTP 732

QY      781  PKETAPTTTPKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKGTATTP 840
DB      781  PKETAPTTTPKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKGTATTP 840

QY      733  PKETAPTTTPKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKGTATTP 792
DB      733  PKETAPTTTPKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKGTATTP 792

QY      841  EPTPKALENSPKFEGVTTTTPAATKPEMTTAKDKTTERDLRTTTPETTTAAAPKOTKETA 900
DB      841  EPTPKALENSPKFEGVTTTTPAATKPEMTTAKDKTTERDLRTTTPETTTAAAPKOTKETA 900

QY      793  EPTPKALENSPKFEGVTTTTPAATKPEMTTAKDKTTERDLRTTTPETTTAAAPKOTKETA 852
DB      793  EPTPKALENSPKFEGVTTTTPAATKPEMTTAKDKTTERDLRTTTPETTTAAAPKOTKETA 852

QY      901  TTTEKTTESKITATTQVTTTQDTPPKITLKITLAPKVTITTKITTTTTEIMNKPE 960
DB      901  TTTEKTTESKITATTQVTTTQDTPPKITLKITLAPKVTITTKITTTTTEIMNKPE 960

QY      853  TTTEKTTESKITATTQVTTTQDTPPKITLKITLAPKVTITTKITTTTTEIMNKPE 912
DB      853  TTTEKTTESKITATTQVTTTQDTPPKITLKITLAPKVTITTKITTTTTEIMNKPE 912

QY      961  ETAKPKDRATNSKATTPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 1020
DB      961  ETAKPKDRATNSKATTPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 1020

QY      913  ETAKPKDRATNSKATTPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 972
DB      913  ETAKPKDRATNSKATTPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 972

QY      1021  NPTSRISAEAMLQTTTRPNQTPNSKLVNPNKSDAGGAGETPHMLLRPHVFPVEVTPDM 1080
DB      1021  NPTSRISAEAMLQTTTRPNQTPNSKLVNPNKSDAGGAGETPHMLLRPHVFPVEVTPDM 1080

QY      973  NPTSRISAEAMLQTTTRPNQTPNSKLVNPNKSDAGGAGETPHMLLRPHVFPVEVTPDM 1032
DB      973  NPTSRISAEAMLQTTTRPNQTPNSKLVNPNKSDAGGAGETPHMLLRPHVFPVEVTPDM 1032

QY      1081  DYLPRVFNQGIINPMLS 1098
DB      1033  DYLPRVFNQGIINPMLS 1050

RESULT 15
US-07-757-022B-58
; Sequence 58, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```

, APPLICATION NUMBER: US 07/757,022B
, FILING DATE: 19910910
, CLASSIFICATION: 530
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 07/643,502
, FILING DATE: 18-JAN-1991
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 07/546,114
, FILING DATE: 29-JUN-1990
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 07/457,196
, FILING DATE: 29-DEC-1989
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 07/390,901
, FILING DATE: 08-AUG-1989
, ATTORNEY/AGENT INFORMATION:
, NAME: Cseri, Luann
, REGISTRATION NUMBER: 31,822
, REFERENCE/DOCKET NUMBER: GI 5190
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: (617)876-1170
, TELEFAX: (617)876-5851
, INFORMATION FOR SEQ ID NO: 58:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 1049 amino acids
, TYPE: AMINO ACID
, TOPOLOGY: linear
, MOLECULE TYPE: protein
US-07-757-022B-58

```

Query Match	91.4%;	Score 5383.1;	DB 4;	Length 1049;
Best Local Similarity	92.5%;	Pred. NO. 2.9e-156;		
Matches 1016; Conservative	7;	Mismatches 26;	Indels 49;	Gaps 2;

QY	61	KRVCTAELSCGRCFESPERGRECDCAQCKYDKCPDYESFCAVHNPTSPSSKXAP	120
Db	26	-----ELSCGRCFESPERGRECDCAQCKYDKCPDYESFCAEHSVSNQESSSS	79
QY	121	PPSGASQIUKSTTKBSKPPNNKXTKV1BSEBITEKVKONKKNRTKKKPTPKPVPVDEA	180
Db	80	SSSSSSSTIIWKISKSNKSAANREUQK-----KVKONKKNRTKKKPTPKPVPVDEA	131
QY	181	GSGLONGDFKVTTPDTSTTQHNKYSTSPKITTAKPINRPSLPNPSDTSKETSLSVNKET	241
Db	132	GSGLONGDFKVTTPDTSTTQHNKYSTSPKITTAKPINRPSLPNPSDTSKETSLSVNKET	191
QY	241	TVETKETTTTTNKQSTOCKBKTSKAKETOSI EKTSAKDLAPTSKVLAKPTPKAETTTKGP	300
Db	192	TVETKETTTTTNKQSTOCKBKTSKAKETOSI EKTSAKDLAPTSKVLAKPTPKAETTTKGP	251
QY	301	ALTTPEKPTPTPKPEASTTPKEPTPTTIKSAPTTPKEPATTTKSAPTTKPEAPITTK	360
Db	252	ALTTPEKPTPTPKPEASTTPKEPTPTTIKSAPTTPKEPATTTKSAPTTKPEAPITTK	311
QY	361	EPAPTTPKEPATTTKBPATTTKSAPTTKBPAPTTPKKBPATTTKBPATTTKPEPTPT	420
Db	312	EPAPTTPKEPATTTKBPATTTKSAPTTKBPAPTTPKKBPATTTKBPATTTKPEPTPT	371
QY	421	TPKEPATTTKEPATTTKBPAPTAPKXPATTTKBPATTTKBPATTTKGPSTTPKEP	480
Db	372	TPKEPATTTKEPATTTKBPAPTAPKXPATTTKBPATTTKBPATTTKGPSTTPKEP	431
QY	481	APTTTKSAPTTKBPATTTKSAPTTKBPSTTTKBPATTTKBPATTTKBPATTPK	540
Db	432	APTTTKSAPTTKBPATTTKSAPTTKBPSTTTKBPATTTKBPATTTKBPATTPK	491
QY	541	BPAPTTPKEPATTTKKPAPTAPKPEAPTTPKETAPTTPKLTPTTPKLAPTTPEKAP	600
Db	492	BPAPTTPKEPATTTKKPAPTAPKPEAPTTPKETAPTTPKLTPTTPKLAPTTPEKAP	551

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 111.709 Seconds
(without alignments)
3171.696 Million cell updates/sec

Title: SEQ1-C
Perfect score: 5887
Sequence: 1 MAWKTLPIYLLLSVFVIQ.....DMDYLPRVFNQGIINPMWS 1098

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 1356558.seqs, 322682953 residues

Total number of hits satisfying chosen parameters: 1356558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
19: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5872.8	99.8	1140	13	US-10-124-557-104
2	5872.8	99.8	1404	9	US-09-802-207-30
3	5872.8	99.8	1404	11	US-09-897-188-1
4	5872.8	99.8	1404	13	US-10-124-557-2
5	5872.8	99.8	1404	13	US-10-124-557-62
6	5871.9	99.7	1361	13	US-10-124-557-40
7	5647.2	95.9	1354	13	US-10-124-557-48
8	5608.7	95.3	1363	13	US-10-124-557-52
9	5607.9	95.3	1311	13	US-10-124-557-42
10	5607.8	95.3	1320	13	US-10-124-557-46
11	5607.8	95.3	1320	13	US-10-124-557-60
12	5466.9	92.9	1022	13	US-10-124-557-84
13	5399.2	91.7	1314	13	US-10-124-557-50
14	5383.1	91.4	1049	13	US-10-124-557-58

15 5383.1 91.4 1313 13 US-10-124-557-142 Sequence 142, Appl
16 5343.8 90.8 1038 13 US-10-124-557-74 Sequence 74, Appl
17 5343.8 90.8 1270 13 US-10-124-557-44 Sequence 44, Appl
18 5011 85.1 941 13 US-10-124-557-14 Sequence 14, Appl
19 2508.6 42.6 792 9 US-09-802-207-27 Sequence 27, Appl
20 1566.8 26.6 292 16 US-10-468-910-4 Sequence 4, Appl
21 1409.9 23.9 538 14 US-10-038-694-3 Sequence 3, Appl
22 1396.9 23.7 5179 9 US-09-922-217-1068 Sequence 1068, Ap
23 1396.9 23.7 5179 9 US-09-833-263-1068 Sequence 1068, Ap
24 1396.9 23.7 5179 13 US-10-025-380-1068 Sequence 121, Appl
25 1396.9 23.7 5179 16 US-10-734-564-121 Sequence 116, Appl
26 1132.9 19.2 207 13 US-10-124-557-116 Sequence 136, Appl
27 1132.9 19.2 207 13 US-10-124-557-116 Sequence 2, Appl
28 1025.8 17.4 185 16 US-10-468-910-2 Sequence 2, Appl
29 1009.9 17.2 188 14 US-10-038-694-2 Sequence 96, Appl
30 975.8 16.6 220 13 US-10-124-557-96 Sequence 428, Appl
31 968.2 16.4 1460 14 US-10-295-027-428 Sequence 35612, A
32 949.8 16.1 1325 9 US-09-864-761-35612 Sequence 54, Appl
33 926.6 15.7 463 13 US-10-124-557-54 Sequence 108, Appl
34 923.9 15.7 1367 9 US-09-801-368-108 Sequence 94, Appl
35 913.8 15.5 209 13 US-10-124-557-94 Sequence 5784, Ap
36 910.5 15.5 3507 14 US-10-369-493-5784 Sequence 132, Appl
37 908.8 15.4 208 13 US-10-124-557-132 Sequence 70, Appl
38 905.4 15.4 296 13 US-10-124-557-70 Sequence 30, Appl
39 899.7 15.3 231 13 US-10-124-557-30 Sequence 92, Appl
40 893 15.2 204 13 US-10-124-557-92 Sequence 90, Appl
41 886.8 15.1 192 13 US-10-124-557-90 Sequence 88, Appl
42 876.8 14.9 172 13 US-10-124-557-88 Sequence 5013, Ap
43 872.9 14.8 6642 14 US-10-369-493-5013 Sequence 32, Appl
44 872.5 14.8 188 13 US-10-124-557-32 Sequence 106, Appl
45 871 14.8 156 13 US-10-124-557-106

ALIGNMENTS

RESULT 1
US-10-124-557-104
; Sequence 104, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:
 NAME: Cserf, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 876-1170
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 104:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1140 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 104:
 US-10-124-557-104

Query Match 99.8%; Score 5872.8; DB 13; Length 1140;
 Best Local Similarity 96.3%; Pred. No. 1.3e-144;
 Matches 1098; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY	1	MAWKTLPIVLLLLSVFVIQVSSODLSSCAGRCGEGYSRDATCNCYNCHYMCCPDF	60
DB	1	MAWKTLPIVLLLLSVFVIQVSSODLSSCAGRCGEGYSRDATCNCYNCHYMCCPDF	60
QY	61	KRVCTAELSCKRCFESFERGECDCDAQCKKYDKCCPDYSCAEVHNPTSPSSKKAP	120
DB	61	KRVCTAELSCKRCFESFERGECDCDAQCKKYDKCCPDYSCAEVHNPTSPSSKKAP	120
QY	121	PPSGASQITKSTTKSPKPNKKTKKVIIESEITE-----	156
DB	121	PPSGASQITKSTTKSPKPNKKTKKVIIESEITE-----	156
QY	121	PPSGASQITKSTTKSPKPNKKTKKVIIESEITE-----	156
DB	121	PPSGASQITKSTTKSPKPNKKTKKVIIESEITE-----	156
QY	157	-----KVKDKNKNRTKKXPTKPPVVDAGSGLDNGDFKVTTPDST	198
DB	181	KIKSSKNSAANRELQKLVKDKNKNRTKKXPTKPPVVDAGSGLDNGDFKVTTPDST	240
QY	199	TOHNKVSIPKLTAKPNRPSLPNSDTSKETSLVNKEITVETKETTNNKQTSIDG	258
DB	241	TOHNKVSIPKLTAKPNRPSLPNSDTSKETSLVNKEITVETKETTNNKQTSIDG	300
QY	259	KEKTTSAKETOSIEKTSADLAPTSKVLAKPTKPAETTKGPALTTPKEPTTPKPEPAS	318
DB	301	KEKTTSAKETOSIEKTSADLAPTSKVLAKPTKPAETTKGPALTTPKEPTTPKPEPAS	360
QY	319	TPKEPTTPKSAPTTPKEPAPTTTKSAPTTKEPAPTTTKPEPAPTTTKPEPAPTTTKP	378
DB	361	TPKEPTTPKSAPTTPKEPAPTTTKSAPTTKEPAPTTTKPEPAPTTTKPEPAPTTTKP	420
QY	379	APTTTKSAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKP	438
DB	421	APTTTKSAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKP	480
QY	439	EPAPTAPKAPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKSAPTTTKSAPTTTKP	498
DB	481	EPAPTAPKAPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKSAPTTTKSAPTTTKP	540
QY	499	TTKSAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKKX	558
DB	541	TTKSAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKKX	600
QY	559	APTAPKEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTT	618
DB	601	APTAPKEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTT	660
QY	619	PEEPAPTTPKAAANPTKEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKGATPTLKEP	678
DB	661	PEEPAPTTPKAAANPTKEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKGATPTLKEP	720
QY	679	APTTPKPEAPKELAPTTTKETSTTSKAPPTTPKGTAPTTTKPEPAPTTTKPEPAPTTPKG	738
DB	721	APTTPKPEAPKELAPTTTKETSTTSKAPPTTPKGTAPTTTKPEPAPTTTKPEPAPTTPKG	780
QY	739	TAPTTLKEPAPTTPKKAPKELAPTTTKGTSTTSKAPPTTPKGTAPTTTKPEPAPTTPKG	798

DB	781	TAPTTLKEPAPTTPKKAPKELAPTTTKGTSTTSKAPPTTPKGTAPTTTKPEPAPTTPK	840
QY	799	KPAPTTPETPTTSEVSTPTTKETPTTIHKSPDSESTPELSAETPKALENSPKPGVPT	858
DB	841	KPAPTTPETPTTSEVSTPTTKETPTTIHKSPDSESTPELSAETPKALENSPKPGVPT	900
QY	859	TKTPRAATKPEMTTAKDKTTERDLATTPETTTAAAPKWKETATTTTEKTESKITATTQV	918
DB	901	TKTPRAATKPEMTTAKDKTTERDLATTPETTTAAAPKWKETATTTTEKTESKITATTQV	960
QY	919	TSTTTQDTPPTFKITLTKTTTLAPKVTTKTKTTTTEIMNKPETAKPKDRATNSKATTPK	978
DB	961	TSTTTQDTPPTFKITLTKTTTLAPKVTTKTKTTTTEIMNKPETAKPKDRATNSKATTPK	1020
QY	979	PQKPTKAPKPTSTKPKTPMVRKPKTPTPRKMTSTMPELNPTSRIAEAMLOTTTTPN	1038
DB	1021	PQKPTKAPKPTSTKPKTPMVRKPKTPTPRKMTSTMPELNPTSRIAEAMLOTTTTPN	1080
QY	1039	QTPNSKLVENPKSEBAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS	1098
DB	1081	QTPNSKLVENPKSEBAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS	1140

RESULT 2
 US-09-802-207-30
 ; Sequence 30, Application US/09802207
 ; Publication No. US200200868241
 ; GENERAL INFORMATION:
 ; APPLICANT: Warman, Matthew
 ; APPLICANT: Carpten, John
 ; APPLICANT: Trent, Jeffrey
 ; APPLICANT: Marcelino, Jose
 ; TITLE OF INVENTION: Novel Methods and Reagents for the Treatment of Osteoarthritis
 ; FILE REFERENCE: Case-06212
 ; CURRENT APPLICATION NUMBER: US/09/802,207
 ; CURRENT FILING DATE: 2001-08-29
 ; PRIOR APPLICATION NUMBER: 09/619,175
 ; PRIOR FILING DATE: 2000-07-19
 ; PRIOR APPLICATION NUMBER: 60/145,328
 ; PRIOR FILING DATE: 1999-07-23
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: patent in version 3.0
 ; SEQ ID NO 30
 ; LENGTH: 1404
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-802-207-30

Query Match 99.8%; Score 5872.8; DB 9; Length 1404;
 Best Local Similarity 96.3%; Pred. No. 1.6e-144;
 Matches 1098; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY	1	MAWKTLPIVLLLLSVFVIQVSSODLSSCAGRCGEGYSRDATCNCYNCHYMCCPDF	60
DB	1	MAWKTLPIVLLLLSVFVIQVSSODLSSCAGRCGEGYSRDATCNCYNCHYMCCPDF	60
QY	61	KRVCTAELSCKRCFESFERGECDCDAQCKKYDKCCPDYSCAEVHNPTSPSSKKAP	120
DB	61	KRVCTAELSCKRCFESFERGECDCDAQCKKYDKCCPDYSCAEVHNPTSPSSKKAP	120
QY	121	PPSGASQITKSTTKSPKPNKKTKKVIIESEITE-----	156
DB	121	PPSGASQITKSTTKSPKPNKKTKKVIIESEITE-----	156
QY	157	-----KVKDKNKNRTKKXPTKPPVVDAGSGLDNGDFKVTTPDST	198
DB	181	KIKSSKNSAANRELQKLVKDKNKNRTKKXPTKPPVVDAGSGLDNGDFKVTTPDST	240
QY	199	TOHNKVSIPKLTAKPNRPSLPNSDTSKETSLVNKEITVETKETTNNKQTSIDG	258
DB	241	TOHNKVSIPKLTAKPNRPSLPNSDTSKETSLVNKEITVETKETTNNKQTSIDG	300


```

RESULT 3
US-09-897-188-1
; Sequence 1, Application US/09897188
; Publication No. US20040072741A1
; GENERAL INFORMATION:
; APPLICANT: Jay, Gregory D.
; TITLE OF INVENTION: Triborectin Polypeptides and Uses Thereof
; FILE REFERENCE: 21486-026 CIP2
; CURRENT APPLICATION NUMBER: US/09/897,188
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 09/298,970
; PRIOR FILING DATE: 1999-04-23
; PRIOR APPLICATION NUMBER: 09/556,246
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patent In Ver. 2.1

```

Qy	679	APTTPKPAKELAPTTTKSPTSTSDKPAPTTPKGTAPTTPKPAPTTPKPAPTTPKPAPTTPK	738
Db	721	APTTPKPAKELAPTTTKSPTSTSDKPAPTTPKGTAPTTPKPAPTTPKPAPTTPKPAPTTPK	780
Qy	739	TAPTTLKEPAPTTTPKKPAKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKPAPTTPK	798
Db	781	TAPTTLKEPAPTTTPKKPAKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKPAPTTPK	840
Qy	799	KPAPTTPETPPPTTSVSTPTTTKEPTTIHKSPDESTPELSAETPKALENSPKPEGVPT	858
Db	841	KPAPTTPETPPPTTSVSTPTTTKEPTTIHKSPDESTPELSAETPKALENSPKPEGVPT	900
Qy	859	TKTAAANKPEMTTAKDKTIERDLRTPETTTAAAPKMTKETATTTEKTTESKITATTQV	918
Db	901	TKTAAANKPEMTTAKDKTIERDLRTPETTTAAAPKMTKETATTTEKTTESKITATTQV	960

RESULT 5

US-10-124-557-62
 ; Sequence 62, Application US/10124557
 ; Publication No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
 ; Clark, Stephen C.
 ; Jacobs, Kenneth
 ; Hewick, Rodney M.
 ; Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557
 ; FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; APPLICATION NUMBER: US 07/546,114
 ; FILING DATE: 29-JUN-1990
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cseri, Luann
 ; REGISTRATION NUMBER: 31,822
 ; REFERENCE/DOCKET NUMBER: G1 5190
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)876-1170
 ; TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 62:

SEQUENCE CHARACTERISTICS:

LENGTH: 1404 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 62:

US-10-124-557-62

Query Match 99.8%; Score 5872.8; DB 13; Length 1404;
 Best Local Similarity 96.3%; Pred. No. 1.6e-144;
 Matches 1098; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGCGSGYGRDATCNCQYHNECCPDF 60
 DB 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGCGSGYGRDATCNCQYHNECCPDF 60
 QY 61 KRVCVTAELSCGRCFESFERGREGCCDCAOCKKYDKCCPDYESFCAEVHNPTSPSSSKAP 120
 DB 61 KRVCVTAELSCGRCFESFERGREGCCDCAOCKKYDKCCPDYESFCAEVHNPTSPSSSKAP 120
 QY 121 PPAGASQTIKSTYKRSKPNKKTKVIESEITE----- 156
 DB 121 PPAGASQTIKSTYKRSKPNKKTKVIESEITE----- 180
 QY 157 -----KVKDKNKNTKKPKPPVVDVDEAGSLDNGDFKVTTPDTST 198

RESULT 6

US-10-124-557-40

; Sequence 40, Application US/10124557

; Publication No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
 ; Clark, Stephen C.
 ; Jacobs, Kenneth
 ; Hewick, Rodney M.
 ; Gesner, Thomas G.

DB 181 KIKSSKNSAANRELOKLLKVKONKKURTKKKTPKPPVVVDVDEAGSLDNGDFKVTTPDTST 240
 QY 199 TQHNKVSTSPKITTAKPINRPSLPPNSDTSKETSILTUNKETTVETKETTTINKQSTOG 258
 DB 241 TQHNKVSTSPKITTAKPINRPSLPPNSDTSKETSILTUNKETTVETKETTTINKQSTOG 300
 QY 259 KEXTTSAKETQSIEKTSKDLAPTSKVLAKPFPKAEITTKGPAITTPKEPTPTTPKEPAS 318
 DB 301 KEXTTSAKETQSIEKTSKDLAPTSKVLAKPFPKAEITTKGPAITTPKEPTPTTPKEPAS 360
 QY 319 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 378
 DB 361 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 420
 QY 379 APTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 438
 DB 421 APTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 480
 QY 439 EPAPTAPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 498
 DB 481 EPAPTAPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 540
 QY 499 TTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 558
 DB 541 TTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 600
 QY 559 APTAPKPEAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 618
 DB 601 APTAPKPEAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 660
 QY 619 PEEAPTTPKAAAPNTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 678
 DB 661 PEEAPTTPKAAAPNTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 720
 QY 679 APTTPKPAKELAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 738
 DB 721 APTTPKPAKELAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 780
 QY 739 TAPTTLKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 798
 DB 781 TAPTTLKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 840
 QY 799 KPAPTTPETPTTSEVSTPTTKPEPTTIHKSPDESTPELSABPTPKALENSPKPEGVPT 858
 DB 841 KPAPTTPETPTTSEVSTPTTKPEPTTIHKSPDESTPELSABPTPKALENSPKPEGVPT 900
 QY 859 TKTPAATKPEMTTAKDXTTERDLRTTPEITTAAPKMTKETATTTEKTSKITATTTOV 918
 DB 901 TKTPAATKPEMTTAKDXTTERDLRTTPEITTAAPKMTKETATTTEKTSKITATTTOV 960
 QY 919 TSTTTQDTPPKIITLLKTTLLAPKVTTKKTIITTEIMNKPEETAAPKDRATNSKATTEK 978
 DB 961 TSTTTQDTPPKIITLLKTTLLAPKVTTKKTIITTEIMNKPEETAAPKDRATNSKATTEK 1020
 QY 979 PQKPTKAPKPTSTKPKTMPVRVKPKTTPTRKMTSTMPNLNPTSRIAEAMLOTTTRPN 1038
 DB 1021 PQKPTKAPKPTSTKPKTMPVRVKPKTTPTRKMTSTMPNLNPTSRIAEAMLOTTTRPN 1080
 QY 1039 QTPNSKLVENVNPKSEDAGAGETPHMLLRPHVFMPEVTPDNDVLPVPGNGLIINPMLS 1098
 DB 1081 QTPNSKLVENVNPKSEDAGAGETPHMLLRPHVFMPEVTPDNDVLPVPGNGLIINPMLS 1140

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 CambridgePark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/124,557
 FILING DATE: 16-Apr-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Cseir, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851
 INFORMATION FOR SEQ ID NO: 40:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1361 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 40:

US-10-124-557-40

Query Match 99.7%; Score 5871.9; DB 13; Length 1361;
 Best Local Similarity 99.9%; Pred. No. 1.6e-144;
 Matches 1097; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY	1	MAWKTLPYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCQHMECCPDF	60
DB	1	MAWKTLPYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCQHMECCPDF	60
QY	61	KRVCTAELSCGRGCFESPERGECDDAQCKYDKCCPDYVSCAEVHNPTSPSSKKAP	120
DB	61	KRVCTAELSCGRGCFESPERGECDDAQCKYDKCCPDYVSCAEVHNPTSPSSKKAP	120
QY	121	PPSGASQTIKSTTKRSPKPNKKTKKVISEEITEKVNKKNRKTKKTPKPPVVDEA	180
DB	121	PPSGASQTIKSTTKRSPKPNKKTKKVISEEITE-KVNKKNRKTKKTPKPPVVDEA	179
QY	181	GSGLNDGDFKVTPTDSTTQHNKYSTSPKLTAKPINPRESLPNDSKETSLSLVNKET	240
DB	180	GSGLNDGDFKVTPTDSTTQHNKYSTSPKLTAKPINPRESLPNDSKETSLSLVNKET	239
QY	241	TVETKETTNNKQSTDGKSTSAKETQSIKTSKDLAPTSKVLAKTPKAEITTKGP	300
DB	240	TVETKETTNNKQSTDGKSTSAKETQSIKTSKDLAPTSKVLAKTPKAEITTKGP	299
QY	301	ALTTPEKPTPTTKPEASTTKPPTTKISAPTTKEPAPTTTKSAPTTKEPAPTTTK	360
DB	300	ALTTPEKPTPTTKPEASTTKPPTTKISAPTTKEPAPTTTKSAPTTKEPAPTTTK	359
QY	361	EPAPTTKEPAPTTTKPEAPTTTKSAPTTKEPAPTTTKKAPTTKEPAPTTKEPPT	420

DB	360	EPAPTTKEPAPTTTKPEAPTTTKSAPTTKEPAPTTTKKAPTTKEPAPTTKEPPT	419
QY	421	TPKEPAPTTKEPAPTTTKPEAPTTAPKAPAPTTKEPAPTTKEPAPTTKEPPTPKP	480
DB	420	TPKEPAPTTKEPAPTTTKPEAPTTAPKAPAPTTKEPAPTTKEPAPTTKEPPTPKP	479
QY	481	APTTKSAPTTTKPEAPTTTKSAPTTTKPSPTTKPEAPTTKEPAPTTKPKAPTTTK	540
DB	480	APTTKSAPTTTKPEAPTTTKSAPTTTKPSPTTKPEAPTTKEPAPTTKPKAPTTTK	539
QY	541	EPAPTTKEPAPTTTKKAPAPTTKEPAPTTKETAATTPKKLTPTTPEKLAATTPKAP	600
DB	540	EPAPTTKEPAPTTTKKAPAPTTKEPAPTTKETAATTPKKLTPTTPEKLAATTPKAP	599
QY	601	TTPEELAPTTPEBPTPTTPEEPAPTTPKAAAPNTPEPAPTTKEPAPTTKEPAPTTTK	660
DB	600	TTPEELAPTTPEBPTPTTPEEPAPTTPKAAAPNTPEPAPTTKEPAPTTKEPAPTTTK	659
QY	661	ETAPTTKGTAPTTKEPAPTTPKKAPKELAPTTTKETSTTSKAPAPTTPKGTAPTT	720
DB	660	ETAPTTKGTAPTTKEPAPTTPKKAPKELAPTTTKETSTTSKAPAPTTPKGTAPTT	719
QY	721	KEPAPTTKEPAPTTPKGTAPTTKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPPT	780
DB	720	KEPAPTTKEPAPTTPKGTAPTTKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPPT	779
QY	781	PKETAPTTKEPAPTTPKKAPPTTETPPPTTSEVSTPTTKEPTTIHKSPDESTPELSA	840
DB	780	PKETAPTTKEPAPTTPKKAPPTTETPPPTTSEVSTPTTKEPTTIHKSPDESTPELSA	839
QY	841	EPTPKALENSPKPGVPTTKTTPAAATKPEMTTAKOKTTERDLRTTETTTAAAPKETA	900
DB	840	EPTPKALENSPKPGVPTTKTTPAAATKPEMTTAKOKTTERDLRTTETTTAAAPKETA	899
QY	901	TTTEKTTESKITATTTQVTSITTTQTPFKITLTKTTTLPKVTTKTITTEIMNKE	960
DB	900	TTTEKTTESKITATTTQVTSITTTQTPFKITLTKTTTLPKVTTKTITTEIMNKE	959
QY	961	ETAKPKDRATNSKATTPKPKQKTPKAPKPTSTKPKTWPVRVKKPTTTPRMTSTMBEL	1020
DB	960	ETAKPKDRATNSKATTPKPKQKTPKAPKPTSTKPKTWPVRVKKPTTTPRMTSTMBEL	1019
QY	1021	NFTSRIAEAMLOTTTRPNQTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVMPVETPDM	1080
DB	1020	NFTSRIAEAMLOTTTRPNQTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVMPVETPDM	1079
QY	1081	DYLPRVFNQGIINPMLS 1098	
DB	1080	DYLPRVFNQGIINPMLS 1097	

RESULT 7

US-10-124-557-48
 ; Sequence 48, Application US/10124557
 ; Publication No. US20020137894A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Turner, Katherine
 ; Jacobs, Kenneth
 ; Hewick, Rodney M.
 ; Gesner, Thomas G.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 48:
US-10-124-557-48

Query Match          95.9%; Score 5647.2; DB 13; Length 1354;
Best Local Similarity 96.3%; Pred. No. 1.1e-138;
Matches 1057; Conservative 7; Mismatches 26; Indels 8; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQQVSSODLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQQVSSODLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60

QY 61 KVCVTAELSCGRCFESFERGECDCDAOCKYDKCCPDYEFCAEVNPTSPSSKKAP 120
DB 61 KVCVTAELSCGRCFESFERGECDCDAOCKYDKCCPDYEFCAEHSVSENQSSSS 120

QY 121 PPSGASQTIKSTKRSKPPNKKTKVIESEITEKVDKNNKRTKKKPTPKPPVDEA 180
DB 121 SSSSSSTIWKIKSSKNSAANRELQKL-----KVDKNNKRTKKKPTPKPPVDEA 172

QY 181 GSGLONGDEKVTPTDSTTQHNKVSPTSPIITAKPINRPSLPKNSDTSKETSLSVNKET 240
DB 173 GSGLONGDPKVITPTDSTTQHNKVSPTSPIITAKPINRPSLPKNSDTSKETSLSVNKET 232

QY 241 TVETKETTITNNKQSTDGKEKTSKAKETOSIEKTSKADLAPTSKVLAKPTPKAETTTKGP 300
DB 233 TVETKETTITNNKQSTDGKEKTSKAKETOSIEKTSKADLAPTSKVLAKPTPKAETTTKGP 292

QY 301 ALTPKEPTTPPKFPASTPKPEPTTTKSAPTPKBPATTTKSAPTTKSAPTTKBPATTTK 360
DB 293 ALTPKEPTTPPKFPASTPKPEPTTTKSAPTPKBPATTTKSAPTTKSAPTTKBPATTTK 352

QY 361 BPATPKBPATTTKBPATTTKGAPTPKBPATTTKPKKBPATTTKBPATTTKBPATTTK 420
DB 353 BPATPKBPATTTKBPATTTKGAPTPKBPATTTKPKKBPATTTKBPATTTKBPATTTK 412

QY 421 TPKEBPATTTKBPATTTKBPATTAAPKKBPATTTKBPATTTKBPATTTKBPATTTKBP 480
DB 413 TPKEBPATTTKBPATTTKBPATTAAPKKBPATTTKBPATTTKBPATTTKBPATTTKBP 472

QY 481 APPTTKSAPTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP 540
DB 473 APPTTKSAPTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP 532

QY 541 BPATPKBPATTTTKKBPATTAAPKKBPATTTKBPATTTKBPATTTKBPATTTKBPAT 600
DB 541 BPATPKBPATTTTKKBPATTAAPKKBPATTTKBPATTTKBPATTTKBPATTTKBPAT 600

533 EPAPTTPKBPATTTTKKBPATTAAPKKBPATTTPKBPATTTPKKLTPTTPEKLAPTTPEKAP 592
601 TTPEELAPTTPEBPTPTTPEBPAPTTTPEBPAPTTPKAAAPNTPKEPAPTTTPEBPAPTTPK 660
593 TTPEELAPTTPEBPTPTTPEBPAPTTTPEBPAPTTPKAAAPNTPKEPAPTTTPEBPAPTTPK 652
661 ETAPTTTPEKGTAPTTLKEPAPTTTPEBPAPTTTPEBPAPTTTPEBPAPTTTPEBPAPTTT 720
653 ETAPTTTPEKGTAPTTLKEPAPTTTPEBPAPTTTPEBPAPTTTPEBPAPTTTPEBPAPTTT 712
721 KEAPTTTPEBPAPTTTPEKGTAPTTTPEBPAPTTTPEBPAPTTTPEBPAPTTTPEBPAPTTT 780
713 KEAPTTTPEBPAPTTTPEKGTAPTTTPEBPAPTTTPEBPAPTTTPEBPAPTTTPEBPAPTTT 772
781 PKETAPTTTPEBPAPTTTPEKGTAPTTTPEBPAPTTTPEBPAPTTTPEBPAPTTTPEBPAPTT 840
773 PKETAPTTTPEBPAPTTTPEKGTAPTTTPEBPAPTTTPEBPAPTTTPEBPAPTTTPEBPAPTT 832
841 EPTPKALENSPKBPVGVTTKTPAATKPEMTTAKDKTTERDLATTTTETTTAAAPKMTKETA 900
833 EPTPKALENSPKBPVGVTTKTPAATKPEMTTAKDKTTERDLATTTTETTTAAAPKMTKETA 892
901 TTTEKTTESKITATTTQVSTTTQDTPPKITLTKTTLAPKVTTKKTTTTEIMNKPE 960
893 TTTEKTTESKITATTTQVSTTTQDTPPKITLTKTTLAPKVTTKKTTTTEIMNKPE 952
961 ETAKPKDRATNSKATTPKPKOKTPKPKPTSTKPKTMRVRKPKTTPPKMTSTMPEL 1020
953 ETAKPKDRATNSKATTPKPKOKTPKPKPTSTKPKTMRVRKPKTTPPKMTSTMPEL 1012
1021 NPTSRIAEAMLQTTTRENQTPNSKLVENVNPKSEDAGAGETPHMLLRPHVFPVETPDM 1080
1013 NPTSRIAEAMLQTTTRENQTPNSKLVENVNPKSEDAGAGETPHMLLRPHVFPVETPDM 1072
1081 DYLPRVFNQGIINPMLS 1098
1073 DYLPRVFNQGIINPMLS 1090

RESULT 8
US-10-124-557-52
; Sequence 52, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
```


LENGTH: 1311 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-124-557-42

Query Match 95.3%; Score 5607.9; DB 13; Length 1311;
Best Local Similarity 95.4%; Pred. No. 1.1e-137;
Matches 1047; Conservative 0; Mismatches 0; Indels 51; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGCGEGYSRDATCNCYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGCGEGYSRDATCNCYNCQHYMECCPDF 60

QY 61 KRVTAEALCKGRCFESFRGECDCDAQCKKYDKCCPDYEFCAEVAHNPTSPSPSKAP 120
DB 61 KRVTAEALCKGRCFESFRGECDCDAQCKKYDKCCPDYEFCAE----- 106

QY 121 PFGASQTIKSTTKRSPKPNKKTKVIESEITEKVKDKNKKNTKKKTPKPPVDEA 180
DB 107 -----VKDNKKNTKKKTPKPPVDEA 129

QY 181 GSGLDGDFKVITPDTSTTQHNKVSIPKITTAKPINRPSLPNSDTSKETSIVNKET 240
DB 130 GSGLDGDFKVITPDTSTTQHNKVSIPKITTAKPINRPSLPNSDTSKETSIVNKET 189

QY 241 TVETKETTITNTQSDGKEKTSIAKETOSIEKTSKDLAPTSKVLAKPTKPAETTKGP 300
DB 190 TVETKETTITNTQSDGKEKTSIAKETOSIEKTSKDLAPTSKVLAKPTKPAETTKGP 249

QY 301 ALTTKPEPTTPKPEASTTKPEPTTKTSAPTTKPEAPTTKSAPTTKPEAPTTTK 360
DB 250 ALTTKPEPTTPKPEASTTKPEPTTKTSAPTTKPEAPTTKSAPTTKPEAPTTTK 309

QY 361 EPAPTTKPEAPTTKPEAPTTKSAPTTKPEAPTTKPKAPTTKPEAPTTKPEPT 420
DB 310 EPAPTTKPEAPTTKPEAPTTKSAPTTKPEAPTTKPKAPTTKPEAPTTKPEPT 369

QY 421 TKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEPTTKPE 480
DB 370 TKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEPTTKPE 429

QY 481 APTTKSAPTTKPEAPTTKSAPTTKPEPTTKPEAPTTKPEAPTTKPKAPTTPK 540
DB 430 APTTKSAPTTKPEAPTTKSAPTTKPEPTTKPEAPTTKPEAPTTKPKAPTTPK 489

QY 541 EPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAP 600
DB 490 EPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAP 549

QY 601 TTPPELAPTTPEPTTPPEAPTTPKAAAPNTKPEAPTTKPEAPTTKPEAPTTPK 660
DB 550 TTPPELAPTTPEPTTPPEAPTTPKAAAPNTKPEAPTTKPEAPTTKPEAPTTPK 609

QY 661 ETAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAP 720
DB 610 ETAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAP 669

QY 721 KPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAP 780
DB 670 KPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAP 729

QY 781 PKETAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAP 840
DB 730 PKETAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAP 789

QY 841 EPTKALENSKPEGVPTTKTAAPKPMVTAKDKTTERDLRTTPETTTAAPKMTETA 900
DB 790 EPTKALENSKPEGVPTTKTAAPKPMVTAKDKTTERDLRTTPETTTAAPKMTETA 849

QY 901 TTEKTTESKITATTTQVSTTTQDTPFPKITTLLKTTTLAPKVTITTKITTTIMNKPE 960
DB 850 TTEKTTESKITATTTQVSTTTQDTPFPKITTLLKTTTLAPKVTITTKITTTIMNKPE 909

QY 961 ETAKPKDRATNSKATTPK 1020
DB 910 ETAKPKDRATNSKATTPK 969

QY 1021 NPTSRIAEAMLQTTTRNQTNPNSKLVNPKSDDAGAGETPHMLLRPHVFWPEVTPDM 1080
DB 970 NPTSRIAEAMLQTTTRNQTNPNSKLVNPKSDDAGAGETPHMLLRPHVFWPEVTPDM 1029

QY 1081 DYLPRVFNQGIINPMLS 1098
DB 1030 DYLPRVFNQGIINPMLS 1047

RESULT 10
US-10-124-557-46
; Sequence 46, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseiz, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-124-557-46

Query Match 95.3%; Score 5607.8; DB 13; Length 1320;
Best Local Similarity 96.2%; Pred. No. 1.1e-137;
Matches 1056; Conservative 0; Mismatches 0; Indels 42; Gaps 2;

QY 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGCGEGYSRDATCNCYNCQHYMECCPDF 60

Db 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25

QY 61 KRVCIAELSCGRCFESPERGECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120

Db 26 -----EUSCKGRCFESPERGECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 79

QY 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITEKVONKNKRTKKKTKPKPPVVD 180

Db 80 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITE -VKONKNKRTKKKTKPKPPVVD 138

QY 181 GSGLDNGDFKVTTPDSTTQHNKUSTSPKITTAKPINRPSLPNSDTSKETSITVNKET 240

Db 139 GSGLDNGDFKVTTPDSTTQHNKUSTSPKITTAKPINRPSLPNSDTSKETSITVNKET 198

QY 241 TVETKETTITNKQSTDCKEKTTSAKETQSIEKTSADLAPTSKVLAKPTPKAETTTKGP 300

Db 199 TVETKETTITNKQSTDCKEKTTSAKETQSIEKTSADLAPTSKVLAKPTPKAETTTKGP 258

QY 301 ALTTPEKPTPTPKBPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPK 360

Db 259 ALTTPEKPTPTPKBPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPK 318

QY 361 EPAPTTPKEPAPTTPKBPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPK 420

Db 319 EPAPTTPKEPAPTTPKBPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPK 378

QY 421 TPKEPAPTTPKEPAPTTPKBPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPK 480

Db 379 TPKEPAPTTPKEPAPTTPKBPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPK 438

QY 481 APPTTKSAPTTPKEPAPTTPKBPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPK 540

Db 439 APPTTKSAPTTPKEPAPTTPKBPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPK 498

QY 541 EPAPTTPKEPAPTTPKBPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPK 600

Db 499 EPAPTTPKEPAPTTPKBPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPK 558

QY 601 TPPELAPTTPKEPTPTTIKSAPTTPKEPTPTTIKSAPTTPKEPTPTTIKSAPTTPKEPTPTTIK 660

Db 559 TPPELAPTTPKEPTPTTIKSAPTTPKEPTPTTIKSAPTTPKEPTPTTIKSAPTTPKEPTPTTIK 618

QY 661 EPAPTTPKEPTPTTIKSAPTTPKEPTPTTIKSAPTTPKEPTPTTIKSAPTTPKEPTPTTIK 720

Db 619 EPAPTTPKEPTPTTIKSAPTTPKEPTPTTIKSAPTTPKEPTPTTIKSAPTTPKEPTPTTIK 678

QY 721 KEAPTTPKEPTPTTIKSAPTTPKEPTPTTIKSAPTTPKEPTPTTIKSAPTTPKEPTPTTIK 780

Db 679 KEAPTTPKEPTPTTIKSAPTTPKEPTPTTIKSAPTTPKEPTPTTIKSAPTTPKEPTPTTIK 738

QY 781 KEAPTTPKEPTPTTIKSAPTTPKEPTPTTIKSAPTTPKEPTPTTIKSAPTTPKEPTPTTIK 840

Db 739 KEAPTTPKEPTPTTIKSAPTTPKEPTPTTIKSAPTTPKEPTPTTIKSAPTTPKEPTPTTIK 798

QY 841 EPTPKALENSKPEPGVPTTKPAATKPEMTTAKDKTTTERDLRTPTTAAAPKWTETA 900

Db 799 EPTPKALENSKPEPGVPTTKPAATKPEMTTAKDKTTTERDLRTPTTAAAPKWTETA 858

QY 901 TTTEKTTESKITAATTOVSTTQDTPPKITLTKTTLAPKVTITTKTITTTIMNKPE 960

Db 859 TTTEKTTESKITAATTOVSTTQDTPPKITLTKTTLAPKVTITTKTITTTIMNKPE 918

QY 961 ETAKPKDRATNSKATTPKPKQPTKAPKKTSTTKPKXTMPVRKPKTTPTPKMTSTMP 1020

Db 919 ETAKPKDRATNSKATTPKPKQPTKAPKKTSTTKPKXTMPVRKPKTTPTPKMTSTMP 978

QY 1021 NPTSRIAEAMLOTTTRPNQNTNSKLVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDM 1080

Db 979 NPTSRIAEAMLOTTTRPNQNTNSKLVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDM 1038

QY 1081 DYLPRVFNQGIINPMLS 1098

Db 1039 DYLPRVFNQGIINPMLS 1056

RESULT 11

US-10-124-557-60

; Sequence 60, Application US/10124557

; Publication No. US20020137894A1

; GENERAL INFORMATION:

; APPLICANT: Turner, Katherine

; Clark, Stephen C.

; Jacobs, Kenneth

; Hewick, Rodney M.

; Gesner, Thomas G.

; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/124,557

; FILING DATE: 16-Apr-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/643,502

; FILING DATE: 18-JAN-1991

; APPLICATION NUMBER: US 07/546,114

; FILING DATE: 29-JUN-1990

; APPLICATION NUMBER: US 07/457,196

; FILING DATE: 29-DEC-1989

; APPLICATION NUMBER: US 07/390,901

; FILING DATE: 08-AUG-1989

; ATTORNEY/AGENT INFORMATION:

; NAME: Cserr, Luann

; REGISTRATION NUMBER: 31,822

; REFERENCE/DOCKET NUMBER: GI 5190

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 876-1170

; TELEFAX: (617) 876-5851

; INFORMATION FOR SEQ ID NO: 60:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1320 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 60:

US-10-124-557-60

Query Match 95.3%; Score 5607.8; DB 13; Length 1320;

Best Local Similarity 96.2%; Pred. No. 1.1e-137;

Matches 1056; Conservative 0; Mismatches 0; Indels 42; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25

Db 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25

QY 61 KRVCIAELSCGRCFESPERGECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120

Db 26 -----ELSCGRCFESPERGECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 79

QY 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITEKVONKNKRTKKKTKPKPPVVD 180

Db 80 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITE -VKONKNKRTKKKTKPKPPVVD 138

QY 181 GSGLDNGDFKVTTPDSTTQHNKUSTSPKITTAKPINRPSLPNSDTSKETSITVNKET 240

Db 139 GSGLDNGFQVTPDTSTQHNKVSFKIITAKINPRPSLPNSDTSKTSITVNKET 198
Qy 241 TVETKETITTTNKQSTGDKETTSKAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGP 300
Db 199 TVETKETITTTNKQSTGDKETTSKAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGP 258
Qy 301 ALTTKETPTPTPKAPSTTKPEPTPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTK 360
Db 259 ALTTKETPTPTPKAPSTTKPEPTPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTK 318
Qy 361 EPAPPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTK 420
Db 319 EPAPPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTK 378
Qy 421 TPKEAPPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTK 480
Db 379 TPKEAPPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTK 438
Qy 481 APPTTKSAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKK 540
Db 439 APPTTKSAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKK 498
Qy 541 EPAPPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKK 600
Db 499 EPAPPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKK 558
Qy 601 TPPEELAPPTPREPTPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTK 660
Db 559 TPPEELAPPTPREPTPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTK 618
Qy 661 ETAPPTPKGAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKK 720
Db 619 ETAPPTPKGAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKK 678
Qy 721 KEAPPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKK 780
Db 679 KEAPPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKK 738
Qy 781 PKETAPPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKK 840
Db 739 PKETAPPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKK 798
Qy 841 EPTPKALENSPKPGVPTTKTAAKPEMTTAKOKTTERDLRTPTTAAPKMTKETA 900
Db 799 EPTPKALENSPKPGVPTTKTAAKPEMTTAKOKTTERDLRTPTTAAPKMTKETA 858
Qy 901 TTTEKTTESKITATTTQVSTTTQDTTPPKITTLKTTTLAPKVTITTKTITTEIMNKPE 960
Db 859 TTTEKTTESKITATTTQVSTTTQDTTPPKITTLKTTTLAPKVTITTKTITTEIMNKPE 918
Qy 961 ETAKPKDRATNSKATTPKQKPTKAPKPTSTTKPKMTPRVRKPTTTPRKMSTMPBEL 1020
Db 919 ETAKPKDRATNSKATTPKQKPTKAPKPTSTTKPKMTPRVRKPTTTPRKMSTMPBEL 978
Qy 1021 NPTSRIAEAWLOTTTPRPNOTNSKLVNPNKSEDAAGAGETPHMLLRPHVEMPEVTPDM 1080
Db 979 NPTSRIAEAWLOTTTPRPNOTNSKLVNPNKSEDAAGAGETPHMLLRPHVEMPEVTPDM 1038
Qy 1081 DYLPRVFNQGIINPMLS 1098
Db 1039 DYLPRVFNQGIINPMLS 1056

RESULT 12
US-10-124-557-84
; Sequence 84, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.

Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luan
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 1022 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-10-124-557-84

Query Match 92.9%; Score 5466.9; DB 13; Length 1022;
Best Local Similarity 95.1%; Pred. No. 3.7e-134;
Matches 1020; Conservative 0; Mismatches 2; Indels 51; Gaps 2;
Qy 26 DLSSCAGRCGEGYSRDATCNCYHMECCPDFFKVC--ELSCGRCFESFGRGEC 85
Db 1 DLSSCAGRCGEGYSRDATCNCYHMECCPDFFKVC--ELSCGRCFESFGRGEC 58
Qy 86 CDAQCKYDKCCPDYBFCFAEVHNPTSPSSKAPPPSGASQTIKSTKSPKPNKKKT 145
Db 59 CDAQCKYDKCCPDYBFCFAET----- 80
Qy 146 KKVIESIEITKVDNKNKRTKKKTPKPPVDEAGSLDNGDFKVTTPDTSTTQHNKVS 205
Db 81 -----AVKDNKNKRTKKKTPKPPVDEAGSLDNGDFKVTTPDTSTTQHNKVS 129
Qy 206 TSPKITTAKPINRPSLPNSDTSKETSILTNNKETTIVETKETTNNKQSTGKEKTTSA 265
Db 130 TSPKITTAKPINRPSLPNSDTSKETSILTNNKETTIVETKETTNNKQSTGKEKTTSA 189
Qy 266 KETOSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTKKAPSTTKPEPT 325
Db 190 KETOSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTKKAPSTTKPEPT 249
Qy 326 PTTIKSAPPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKK 385
Db 250 PTTIKSAPPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKK 309
Qy 386 APTTKKBPAPPTPKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTK 445

Db 553 TTPBELAPTTPEPTTPEPATTPKQAAAPNTPKBPATTPKEPATTPKBPATTPK 612
Qy 661 ETAPTPKGTAPTTKBPATTPKBPAPKELAPTTTKEPTSTTSKBPATTPKGTATTP 720
Db 613 ETAPTPKGTAPTTKBPATTPKBPAPKELAPTTTKEPTSTTSKBPATTPKGTATTP 672
Qy 721 KEPATTPKBPATTPKGTATTPKBPATTPKBPAPKELAPTTTKEPTSTTSKBPATTP 780
Db 673 KEPATTPKBPATTPKGTATTPKBPATTPKBPAPKELAPTTTKEPTSTTSKBPATTP 732
Qy 781 PKETAPTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 840
Db 733 PKETAPTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 792
Qy 841 EPTPKALENSKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 900
Db 793 EPTPKALENSKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 852
Qy 901 TTEKTTESKITATTTQVTSITTTQVTSITTTQVTSITTTQVTSITTTQVTSITTT 960
Db 853 TTEKTTESKITATTTQVTSITTTQVTSITTTQVTSITTTQVTSITTTQVTSITTT 912
Qy 961 ETAPKDRATNSKATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 1020
Db 913 ETAPKDRATNSKATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 972
Qy 1021 NPTSRIAEAMLTQTTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 1080
Db 973 NPTSRIAEAMLTQTTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 1032
Qy 1081 DYLPRVNPQGIINPMLS 1098
Db 1033 DYLPRVNPQGIINPMLS 1050

RESULT 14

US-10-124-557-58

; Sequence 58, Application US/10124557

; Publication NO. US20020137894A1

; GENERAL INFORMATION:

; APPLICANT: Turner, Katherine

; Jacobs, Kenneth

; Hewick, Rodney M.

; Gesner, Thomas G.

; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESS: Genetics Institute, Inc.

; STREET: 87 Cambridgepark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; Zip: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/124,557

; FILING DATE: 16-APR-2002

; CLASSIFICATION: <Unknown>

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 07/643,502

; FILING DATE: 18-JAN-1991

; APPLICATION NUMBER: US 07/546,114

; FILING DATE: 29-JUN-1990

; APPLICATION NUMBER: US 07/457,196

; FILING DATE: 29-DEC-1989

; APPLICATION NUMBER: US 07/390,901

; FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:

LENGTH: 1049 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 58:

US-10-124-557-58

Query Match 91.4%; Score 5383.1; DB 13; Length 1049;

Best Local Similarity 92.5%; Pred. No. 5.7e-132;

Matches 1046; Conservative 7; Mismatches 26; Indels 49; Gaps 2;

Qy 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25

Db 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25

Qy 61 KRYCTAELSCGRCFSEFERGECDDAOCKYDKCCPDYEGFCABVHNPTSPSSSKAP 120

Db 26 -----ELSCGRCFSEFERGECDDAOCKYDKCCPDYEGFCABVHNPTSPSSSKAP 79

Qy 121 PPSGASQTKITTKRSPKPPKTKKVIKVESEITEKVDKDKKTKKTKKTKKTKKTKK 180

Db 80 SSSSSSTIWKIKSSKNSAANRELQKLL-----KVDKDKKTKKTKKTKKTKKTKK 131

Qy 181 GSGLDNGDFKVTTPDTSTTHNKKVSTSPKLTAKPINRPSLPNSDTSKETSITVKNKT 240

Db 132 GSGLDNGDFKVTTPDTSTTHNKKVSTSPKLTAKPINRPSLPNSDTSKETSITVKNKT 191

Qy 241 TVETKETTITNKOTSTDGKEKTSKETSISAKDLAPTSSKVLAKTPPKAETTKGP 300

Db 192 TVETKETTITNKOTSTDGKEKTSKETSISAKDLAPTSSKVLAKTPPKAETTKGP 251

Qy 301 ALATPKETPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 360

Db 252 ALATPKETPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 311

Qy 361 EPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 420

Db 312 EPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 371

Qy 421 TPKEPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 480

Db 372 TPKEPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 431

Qy 481 APITTKSAPTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 540

Db 432 APITTKSAPTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 491

Qy 541 EPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 600

Db 492 EPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 551

Qy 601 TTPBELAPTTPEEPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 660

Db 552 TTPBELAPTTPEEPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 611

Qy 661 ETAPTPKGTAPTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 720

Db 612 ETAPTPKGTAPTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 671

Qy 721 KEPATTPKBPATTPKGTAPTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 780

Db 672 KEPATTPKBPATTPKGTAPTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 731

Qy 781 PKETAPTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 840

Db	912	ETAKPKDRATNSKATTPKQKPTKAPKKPTSTKKPKTTPRVRKPKTTPRKOVTSMPEL	971
Qy	1021	NPTSRIAEAMLQTTTRENQTPNSKLVENVNPKSEDAGAGETPHMLLRPHVMPVETPDM	1080
Db	972	NPTSRIAEAMLQTTTRENQTPNSKLVENVNPKSEDAGAGETPHMLLRPHVMPVETPDM	1031
Qy	1081	DYLPRVFNQGIINPMLS	1098
Db	1032	DYLPRVFNQGIINPMLS	1049

Search completed: October 13, 2004, 11:52:51
Job time : 117.709 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw.model

Run on: October 13, 2004, 11:23:49 ; Search time 120.725 Seconds
(without alignments)
5233.063 Million cell updates/sec

Title: SEQ1-C
Perfect score: 587
Sequence: 1 NAWKTLPIYLLLLSVFIQ.....DMDYLPRVNCGIINPMLS 1098

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5872.8	99.8	1404	2	Q92954 homo sapien
2	5863.8	99.6	1404	2	Q9BX49
3	3204.7	54.4	933	2	Q6ZM25
4	3204.7	54.4	933	2	BAD18580
5	2731.5	46.4	1054	2	Q9JW99
6	1396.9	23.7	5179	1	MUC2 HUMAN
7	1327.5	22.5	1761	2	Q9VR49
8	1327.5	22.5	1761	2	Q7KTF6
9	1266.3	21.5	3150	2	AAS4673
10	1266.3	21.5	3150	2	Q6SSE6
11	1205.1	20.5	3409	2	AAS07044
12	1205.1	20.5	3409	2	SLPI_CLOTM
13	1132	19.2	1684	2	Q6SSE8
14	1124.7	19.1	3889	2	AAS07042
15	1124.7	19.1	3889	2	Q8WQ4
16	1115.3	18.9	1349	2	Q8IR51
17	1108.1	18.8	3432	2	Q8IR52
18	1108.1	18.8	3432	2	Q76894
19	1100.7	18.7	1795	2	Q9N457
20	1065.6	18.1	1079	2	Q9N457
21	1054.7	17.9	23015	2	Q81Q18
22	1054.7	17.9	23015	2	AAAI0358
23	1049.8	17.8	9234	2	Q7KTP5
24	1049.8	17.8	9234	2	AAAI0531
25	1047.2	17.8	2284	2	Q9VPG1
26	1042.9	17.7	1607	2	Q8H6Q5
27	1042.9	17.7	1607	2	AAP74661
28	1032.6	17.5	5703	1	MUSB_HUMAN
29	1028.4	17.5	972	2	Q7QKK7
30	1020.9	17.3	1489	2	Q96449
31	1016.3	17.3	34350	2	Q8WZ42

32	1010.7	17.2	7962	2	Q10465
33	1008.3	17.1	1274	2	Q20007
34	1007.4	17.1	10578	2	Q8ISF5
35	1005.4	17.1	18519	2	Q8ISF6
36	1005.4	17.1	18534	2	Q8ISF7
37	996.9	16.9	1480	2	Q9LIE8
38	996.8	16.9	2187	2	P70670
39	975	16.6	2768	2	Q9VC00
40	969.5	16.5	1458	2	Q757N5
41	969.5	16.5	1458	2	AAS2662
42	965.9	16.4	2112	2	Q9VEL9
43	964.4	16.4	1121	2	Q72884
44	960.4	16.3	4498	2	Q9W223
45	960.2	16.3	2042	2	Q767L8

ALIGNMENTS

RESULT 1

Q92954	PRELIMINARY;	PRT; 1404 AA.
AC	Q92954;	
DT	01-FEB-1997 (Tremblrel_02, Created)	
DT	01-FEB-1997 (Tremblrel_02, Last sequence update)	
DT	01-MAR-2004 (Tremblrel_26, Last annotation update)	
DE	Megakaryocyte stimulating factor.	
OS	Homo sapiens (Human)	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,	
RA	Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,	
RA	Ferez C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,	
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.;	
RT	"Purification, Biochemical Characterization, and Cloning of a Novel	
RT	Megakaryocyte Stimulating Factor that has Megakaryocyte Colony	
RT	Stimulating Activity."	
RL	Blood 78:279-279(1991).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RA	Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,	
RA	Fitzgerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,	
RA	Jacobs K., Turner K.;	
RT	"A Comparison of Vitronectin and Megakaryocyte Stimulating Factor."	
RT	(In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,	
RL	Mosher D.F. (eds.);	
RL	BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS., pp.45-52, Elsevier	
RL	Science Publishers B.V. (1993).	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,	
RA	Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,	
RA	Ferez C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,	
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.;	
RL	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; U70136; AA09089.1; -.	
DR	HSSP; P04004; IOCO.	
DR	Genew; HGNC:9364; PRG4.	
DR	GO; GO:0008283; P:cell proliferation; TAS.	
DR	InterPro; IPR001212; Somatomedin_B.	
DR	InterPro; IPR001212; Somatomedin_B.	
DR	Pfam; PF00045; Hemopexin; 2.	
DR	Pfam; PF00045; Hemopexin; 2.	
DR	PRINTS; PRO0022; SOMATOMEDIN_B; 2.	
DR	SMART; SM00201; SO; 2.	
DR	SMART; SM00201; SO; 2.	
DR	PROSITE; PS00024; HEMOPEXIN; UNKNOWN 1.	
DR	PROSITE; PS00524; SOMATOMEDIN_B; 2.	
DR	PROSITE; PS00524; SOMATOMEDIN_B; 2.	
DR	SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;	
SQ		

Query Match									
Best Local Similarity 99.8%; Score 5872.8; DB 2; Length 1404;									
Matches 1096; Conservative 0; Mismatches 0; Indels 42; Gaps 1;									
QY	1	MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF	60						
DB	1	MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF	60						
QY	61	KRVCTAELSCGRCFESFERGECDCDAQCKYDKCCPDYEFCAEVNPTSPSSKKAP	120						
DB	61	KRVCTAELSCGRCFESFERGECDCDAQCKYDKCCPDYEFCAEVNPTSPSSKKAP	120						
QY	121	PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITE-----	156						
DB	121	PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITE-----	156						
QY	157	-----KVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST	180						
DB	157	-----KVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST	180						
QY	181	KIKSSNSAANRELOKLVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST	240						
DB	181	KIKSSNSAANRELOKLVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST	240						
QY	199	TOHNVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTETKTTNKQSTDG	258						
DB	199	TOHNVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTETKTTNKQSTDG	258						
QY	241	TOHNVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTETKTTNKQSTDG	300						
DB	241	TOHNVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTETKTTNKQSTDG	300						
QY	259	KEKTTSAKETOSIEKTSKOLAPTSKVLAKPTPKAETTTKGPALTTPKKEPAP	318						
DB	301	KEKTTSAKETOSIEKTSKOLAPTSKVLAKPTPKAETTTKGPALTTPKKEPAP	360						
QY	319	TTPKETPTTIKSAPTTPKEPAPTTTKSAPTTPKKEPAPTTTKKEPAPTTTKKEP	378						
DB	361	TTPKETPTTIKSAPTTPKEPAPTTTKSAPTTPKKEPAPTTTKKEPAPTTTKKEP	420						
QY	379	APTITKSAPTTPKEPAPTTTKKAPTTPKKEPAPTTPKKEPAPTTTKKEPAPTTK	438						
DB	421	APTITKSAPTTPKEPAPTTTKKAPTTPKKAPTTPKKAPTTPKKAPTTPKKAPTTK	480						

DB	1021	PQKTKAPKPKTSTTKPKTTPRVKPKTTPTRKMTSTPELNPTSRIAEAMLOTTREN	1080						
QY	1039	QTPNSKLVEYNPKSDEAGGAGETPHMLRPHVFMPEVTPDMYLPVRNQGIIINPMLS	1098						
DB	1081	QTPNSKLVEYNPKSDEAGGAGETPHMLRPHVFMPEVTPDMYLPVRNQGIIINPMLS	1140						
RESULT 2									
Q9BX49 PRELIMINARY; PRT; 1404 AA.									
ID	Q9BX49	AC	Q9BX49						
DT	01-JUN-2001	(T-EMBLrel. 17, Created)							
DT	01-JUN-2001	(T-EMBLrel. 17, Last sequence update)							
DT	01-MAR-2004	(T-EMBLrel. 26, Last annotation update)							
DE	BGI74L6.2	(MSF: megakaryocyte stimulating factor)							
GN	Name=BG174L6.2;								
OS	Homo sapiens (Human)								
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
OX	NCBI_TaxID=9606;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RA	Wray P.;								
RL	Submitted (JUL-2000) to the EMBL/GenBank/DDAJ databases.								
RL	EMBL; AL133553; CAC36090.1; --								
DR	HGSP; P04004; LOC0.								
DR	InterPro; IPR000595; Hemopexin.								
DR	InterPro; IPR001212; Somatomedin_B.								
DR	Pfam; PF00045; Hemopexin; 2.								
DR	Pfam; PF01033; Somatomedin_B; 2.								
DR	PRINTS; PR00022; SOMATOMEDINB.								
DR	SMART; SM00120; HX; 2.								
DR	SMART; SM00201; SO; 2.								
DR	PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.								
DR	PROSITE; PS00024; SOMATOMEDIN_B; 2.								
SQ	SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDBE5 CRC64;								
Query Match 99.6%; Score 5863.8; DB 2; Length 1404;									
Best Local Similarity 96.1%; Pred. No. 2.4e-104;									
Matches 1096; Conservative 0; Mismatches 2; Indels 42; Gaps 1;									
QY	1	MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF	60						
DB	1	MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF	60						
QY	61	KRVCTAELSCGRCFESFERGECDCDAQCKYDKCCPDYEFCAEVNPTSPSSKKAP	120						
DB	61	KRVCTAELSCGRCFESFERGECDCDAQCKYDKCCPDYEFCAEVNPTSPSSKKAP	120						
QY	121	PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITE-----	156						
DB	121	PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITE-----	156						
QY	157	-----KVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST	180						
DB	157	-----KVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST	180						
QY	181	KIKSSNSAANRELOKLVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST	240						
DB	181	KIKSSNSAANRELOKLVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST	240						
QY	199	TOHNVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTETKTTNKQSTDG	258						
DB	199	TOHNVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTETKTTNKQSTDG	258						
QY	241	TOHNVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTETKTTNKQSTDG	300						
DB	241	TOHNVSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTETKTTNKQSTDG	300						
QY	259	KEKTTSAKETOSIEKTSKOLAPTSKVLAKPTPKAETTTKGPALTTPKKEPAP	318						
DB	301	KEKTTSAKETOSIEKTSKOLAPTSKVLAKPTPKAETTTKGPALTTPKKEPAP	360						
QY	319	TTPKETPTTIKSAPTTPKEPAPTTTKSAPTTPKKEPAPTTTKKEPAPTTTKKEP	378						
DB	361	TTPKETPTTIKSAPTTPKEPAPTTTKSAPTTPKKEPAPTTTKKEPAPTTTKKEP	420						
QY	379	APTITKSAPTTPKEPAPTTTKKAPTTPKKEPAPTTPKKEPAPTTTKKEPAPTTK	438						
DB	421	APTITKSAPTTPKEPAPTTTKKAPTTPKKAPTTPKKAPTTPKKAPTTPKKAPTTK	480						


```

RL Cytogenet. Cell Genet. 90:291-297 (2000).
DR EMBL; AB034730; BAA92310.1; -.
DR HSSP; P04004; I0C0.
DR MGD; MGI:1891344; Pr94.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemopexin; 2.
DR Pfam; PF01033; Somatomedin_B; 2.
DR PRINTS; PR00022; SOMATOMEDIN_B.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
SQ SEQUENCE 1054 AA; 115991 MW; 4FC64BFA42283235 CRC64;

Query Match 46.4%; Score 2731.5; DB 2; Length 1054;
Best Local Similarity 47.1%; Pred. No. 9.3e-45;
Matches 545; Conservative 43; Mismatches 143; Indels 425; Gaps 21;

QY 1 MAWKTLPIYILLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMCCPDF 60
Db 1 MGWKILPVCLSLLPVLLIQVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMCCPDF 60

QY 61 KRYCTAELCKGRCFSPERGECDDAOKKYDKCCPDYSEFCABVHNPTSPPSKKAP 120
Db 61 KRYCSPSELCKGRCFSPERGECDDSQCKQYKCCADYDFCEVHNSTSPSSKTAP 119

QY 121 PPSGASQTIKSTTKRKPKNKKTKKVJESSEITEK----- 157
Db 120 TPAGASDTIKSTTKRSPKFT-TRTIKVVESELTEHSDSENQSSSSSSSSSTIRKI 178

QY 158 -----VKONKNRTKKKTPPKPPVVDVDEAGSLDNGDFKVT--TDTSTT 199
Db 179 KSSKSNANRELQNPVNVKNTKPKKENPEPPAVDEAGSLDNGDFKLTTPPDPTT 238

QY 200 QENKYSTPKITTAKEINPRPSLPNSDTSKTSITVNETVETTKNTKQSTOCK 259
Db 239 PISKVAISPKTTAAKPVTPKSLAPNSETSKASLASNETIVETKTTATNKQSA-SK 297

QY 260 EKTTSAKETQSIEKTSKAKLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPKEPAST 319
Db 298 KKTTSVKETRSABKTSKDVETPS----- 321

QY 320 TPKEPTPTTKSAPTTKPEPAPTTTKSAPTTKPEPAPTTKPEPAPTT-----PKBPA 371
Db 322 -----TTPKNSAPTTTKKPVTTTKESKFLPQBPPE 352

QY 372 PTTTKEPAPTTTKSAPTTKPEPAPTTKPKPAPTTTKEPAPTTKPEPAPTTKE 431
Db 353 PTTAKEPPTTKKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKK 412

QY 432 PAPTTPKEPAPTAAPKAPAPTTKPEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 491
Db 413 PEPTTPKEPAPTTPKEPPTTKKEPPTTKKEPPTTKKEPPTTKKEPPTTKPE 461

QY 492 TKEPAPTTTKSAPTTKPEPPTTKPEPAPTTTKEPAPTTKPEPAPTTKPEPAPTT 551
Db 462 -----BPTTKPEPPTTKPEPPTTKPEPPTTKPEPPTTKPEPPTTKPE 510

QY 552 PTTTKEPAPTAAPKAPAPTTKPEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 611
Db 511 PT----- 512

QY 612 EBPPTTPPEPAPTTKAAAPNTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTT 671
Db 513 -----TPKEPPTTKPEPPTTKPEPPTTKPEPPTTKPEPPTTKPEPPT 538

QY 672 PTTTKEPAPTTKPKAPKAPLAPTTTKEPSTTSKDPAPTTKGPAPTTPKEPAPTT 731
Db 539 ----KEPPTTKPKP-----BPTTKPEPPTTKPEP 565

QY 732 ATTPKGTAPTLKEPAPTTKPKAPKAPLAPTTTKEPSTTSKDPAPTTKPEPAPTT 791
Db 731 ----- 791

```

```

Db 566 EPTTP-----KEPPTTKPEP-----TTRKEPPTTKPEPPTTKPE 604
QY 792 PAPTTPKKAPATTPTETPTTSEVSTPTTKPTTIHKSDPESTPELSAETPKALENSP 851
Db 605 PEPTTPKKPEPTT----- 617
QY 852 KPGVPTTKTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETATTTEKTESKI 911
Db 618 -----TSFKT----- 622
QY 912 TATTQVSTTTQDTTPTFKITLTKTTLAPKVVTTTKTITTTTEIMNKPEETAKPKDRATN 971
Db 623 -----TTLKATTLAPKVTAPAE-----ETQNKPEETTPASESDSD 657
QY 972 SKAT-----TPKP-QKPTKAPKPTSTKPKTPRVRKPKTTPPRKWTSTMPELNP 1022
Db 658 SKTLKPKQPKAPKPKTKPKAPKPTSTKPKT-PATRKPKTTPAPLTKTTSATPELNT 716
QY 1023 TSRIAEAMLQTTTTPNQTTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPMDY 1082
Db 717 TP--LEVMLPTTTPKQTPENPETAENVNPDHEDADGGEKFP-LIPGPPVLPFTAIPGDL 773
QY 1083 LPRVPCGIIINPMLS 1098
Db 774 LAGRLNRGININPMPS 789

RESULT 6
MUC2_HUMAN
ID MUC2_HUMAN STANDARD; PRT; 5179 AA.
AC Q02817; Q14878;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Mucin 2 precursor (Intestinal mucin 2).
GN Name=MUC2; Synonyms=SMUC;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor.";
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region.";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE=91358717; PubMed=1885763;
RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism.";
RL J. Clin. Invest. 88:1005-1013(1991).
CC -!- FUNCTION: Coats the epithelia of the intestines, airways, and
CC other mucous membrane-containing organs. Thought to provide a
CC protective, lubricating barrier against particles and infectious
CC agents at mucosal surfaces.
CC -!- SUBUNIT: Multimeric.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Colon, small intestine, colonic tumors,

```

CC bronchus, cervix and gall bladder.
 CC -!- PTM: All cysteine residues are involved in intrachain or
 CC intrachain disulfide bonds (By similarity).
 CC -!- POLYMORPHISM: The number of repeats is highly polymorphic and
 CC varies among different alleles.
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
 CC -!- SIMILARITY: Contains 1 TIL (Typein inhibitory-like) domain.
 CC -!- SIMILARITY: Contains 2 WFCC domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L21998; AAB95295.1; -
 CC EMBL; M74027; AAA59875.1; -
 CC EMBL; M94131; AAA59163.1; -
 CC EMBL; M94132; AAA59164.1; -
 CC F01; A49563; A43932.
 CC Genew; HGNC:7512; MUC2.
 CC MIM; 158370; -
 CC InterPro: IPR002919; Cysrich TIL.
 CC InterPro: IPR006208; Cys knot.
 CC InterPro: IPR006207; Cys_knot_C.
 CC InterPro: IPR006209; EGF_like.
 CC InterPro: IPR001007; WFC_C.
 CC InterPro: IPR001846; WFC_D.
 CC Pfam; PF00007; Cys_knot; 1.
 CC Pfam; PF01826; TIL; 1.
 CC Pfam; PF00093; WVC; 1.
 CC Pfam; PF00094; WVD; 1.
 CC SMART; SM00041; CT; 1.
 CC SMART; SM00214; WVC; 2.
 CC SMART; SM00216; WVD; 4.
 CC PROSITE; PS01185; CTCK_1; 1.
 CC PROSITE; PS01225; CTCK_2; 1.
 CC PROSITE; PS00022; EGF_1; UNKNOWN_1.
 CC PROSITE; PS01208; WFC_1; 2.
 CC PROSITE; PS0184; WFC_2; 2.
 KW Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 20 Potential.
 FT CHAIN 21 5179 Mucin 2.
 FT DOMAIN 1401 1747 APPROXIMATE REPEATS.
 FT REPEAT 1401 1416 1.
 FT REPEAT 1417 1432 2.
 FT REPEAT 1433 1448 3.
 FT REPEAT 1449 1464 4.
 FT REPEAT 1465 1471 5.
 FT REPEAT 1472 1478 6.
 FT REPEAT 1479 1494 7A.
 FT REPEAT 1495 1517 7B.
 FT REPEAT 1518 1533 8A.
 FT REPEAT 1534 1556 8B.
 FT REPEAT 1557 1572 9A.
 FT REPEAT 1573 1596 9B.
 FT REPEAT 1597 1612 10A.
 FT REPEAT 1613 1635 10B.
 FT REPEAT 1636 1651 11A.
 FT REPEAT 1652 1675 11B.
 FT REPEAT 1676 1693 12.
 FT REPEAT 1684 1699 13.
 FT REPEAT 1700 1715 14.
 FT REPEAT 1716 1731 15.
 FT REPEAT 1732 1747 16.
 FT DOMAIN 4815 4886 WFC 1.
 FT DOMAIN 4924 4991 WFC 2.
 FT DOMAIN 5075 5160 CTCK.
 FT DISULFID 5075 5122 By similarity.
 FT DISULFID 5089 5136 By similarity.
 FT DISULFID 5089 5152 By similarity.

FT DISULFID 5102 5154 By similarity.
 FT DISULFID 5159 5159 By similarity.
 FT CARBOHYD 163 163 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 423 423 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 670 670 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 770 770 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 894 894 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1139 1139 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1154 1154 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1215 1215 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1230 1230 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1246 1246 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1787 1787 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1820 1820 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4339 4339 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4351 4351 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4362 4362 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4373 4373 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4422 4422 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4438 4438 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4502 4502 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4615 4615 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4627 4627 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4752 4752 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4787 4787 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4881 4881 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4888 4888 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4955 4955 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 4970 4970 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 5019 5019 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 5038 5038 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 5069 5069 N-linked (GlcNAc...) (Potential).
 FT CONFLICT 1351 1351 H -> L (in Ref. 3).
 FT CONFLICT 1412 1412 T -> S (in Ref. 3).
 FT CONFLICT 1449 1449 L -> P (in Ref. 3).
 FT CONFLICT 1504 1504 M -> T (in Ref. 3).
 FT CONFLICT 4192 4192 G -> S (in Ref. 2).
 SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;
 Query Match 23.7%; Score 1396.9; DB 1; Length 5179;
 Best Local Similarity 20.1%; Pred. No. 1.9e-18;
 Matches 449; Conservative 85; Mismatches 494; Indels 1211; Gaps 71;
 QY 6 LPVYLLLSVFVIQQVSSQDLSSCGRG-----EG-----YSR 40
 DB 502 LQVQLAPVWQLFVTLDAQS--GQVGLCGNFGLEGDFKTASGLVEATGAGFANTWKA 559
 QY 41 DATCN-----CDY---- 48
 DB 560 QSTCHDKLDLDDPCSLINIESANYAEHWCSSLKKTETPGRCHSAVDPAEYKRYKDYTC 619
 QY 49 NCOHYMEC-----CPDFKRYCTA-----ELSK 71
 DB 620 NQNNEDCLCAALSRYARACTAKGVMLGWREHVCNKGVCNFSQVFLNYLTTCQOTCR 679
 QY 72 -----GRCFESP-----ERGR-----ECDC-----DAQCKYDK 95
 DB 680 SLSEADSHCLEGAPVVDGCGCPDHTFLDEKRCVPLAKSCYHRGLYLAGDVVQEEER 739
 QY 96 C-----YESFCAEVHN----- 96
 DB 740 CVCRDRLHCRQIRLIGQSCTAPKIHMDCSNLTALATSKPRALSCOTLAAGYHTECVSG 799
 QY 97 --CPD-----YESFCAEVHN----- 109
 DB 800 CVCPDGLMDGRGCGVVEKECPCVHNNDLYSSGAKIKVDCNTCTCKRGWVCTQAVCHT 859
 QY 110 ----- 109
 DB 860 CSYVSGSHYITFDGKYVDFDGHGCVAVQDYCQNSLGSLSIITENVPGTGTGVTCSKA 919
 QY 110 ----- 109

Db 920 IKIFMGTELEKLEDRVVIQRDEGHVAYTTREVCGQYLVBESSGCIIVINDKRTVTFIK 979
QY 110 -----PTSPSSKKAAPPSS- 123
Db 980 LAPSYKGTVCGLCGNFDRHSNNDFTTRDHVWVSSELDGNSWKEAPCTPDVSTNPEPCSL 1039
QY 124 -----GASQTIKST-----TKRSPK- 139
Db 1040 NPHRRSWAEKQCSILKSVFSICHKVDKPKPFYACVHDSKSCDGTGGDCBCEFCSAVYA 1099
QY 140 -----PNKKKTKVIESBEITEKVDNKKNRKKK-----PTPKPP----- 139
Db 1100 QECTKEGACVFWRTPLDCIPFCDYNNPHECEWHYBPCGNRSFETCORTINGIHSNISVSY 1159
QY 140 -----PNKKKTKVIESBEITEKVDNKKNRKKK-----PTPKPP----- 175
Db 1160 LEGYPCRCADRIYE-----EDLKKCVTADKCGCVEDTHYPPGASVTEBCKSC 1211
QY 176 -----VDEAGSGL-----DNG----- 187
Db 1212 VCTNSSQVCRPEGKILNQDGFACYWEICGPNGTVEKHFNCISITRPSLTITFTI 1271
QY 188 -----DFKVTPTDSTQHNNKVSPPK----- 209
Db 1272 TUPPTPTSTTTTTPTSTSTVLSTPKLCLWSDWINEHPSGSDGDRPEFDGVC 1331
QY 210 -----TPSPDISITTTTTPTSTSTVLSTPKLCLWSDWINEHPSGSDGDRPEFDGVC 209
Db 1332 APEDIECRSVKPHLSLEOHGQKQVQDVSQVGRICKNEDQFNGPGLCYDYKIRVNCWP 1391
QY 210 -----INTAKPINRPSLPNSDTSKETSITVKNKETTVEKETTITNNKQSTDGKSKTISA 265
Db 1392 MDKCIITPSPPPTTTPSPPTTTLPTTTPSPPTT-----TTTTPPTTTPSPPTT 1446
QY 266 KETQSEKTSKADLAPTSKVLAKPTKAEITTKGPAITTKPEPT-----PTPKPASTT 320
Db 1447 TP-----TPSPDISITTTTTPTSTSTVLSTPKLCLWSDWINEHPSGSDGDRPEFDGVC 1490
QY 321 PKEPTPTTIKSAF-TTPKEP-----APTITKSAF-TTPKEPAPTITKPEAPTTPKEP- 370
Db 1491 TTTTPPTTTPSPPTTTPSPPTTTLPTTTPSPPTTTLPTTTPSPPTTTPSPPTTTPSPPT 1550
QY 371 -----APTITKSAF-TTPKEPAPTITKPEAPTTPKEPAPTTPKEAPTTPKEAPTTP 422
Db 1551 TSTTLPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTT 1610
QY 423 KEPAAPTTPKEAPTTPKEP-----APTAPKPAAPTTPKEAPTTPKEAPTTPKEAPT 475
Db 1611 TPTPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSP 1670
QY 476 TPKPEAPTTPKSAAPTTPKEAPTTPKSAF-----TTPKEPSPPTTPKSAAPTTPKEAPTTPK 532
Db 1671 TTTTPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTT 1729
QY 533 KPAAPTTPKEAPTTPKEAPTTPKPAAPTTPKEAPTTP-----KETAPTTPKLTTP--- 584
Db 1730 TSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTP 1787
QY 585 -----TPKEKAPAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPT 584
Db 1788 WTGWLDSGKPNFKPGDTELIGDVGCGWAANISCRATMYPDVPIGQLGQVWCDVSG 1847
QY 585 -----TPKEKAPAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPTTPKEAPT 612
Db 1848 LICKEDQKGGVIMAFCLNVEINVQCECVTQPTTMTTNTTNTTNTTNTTNTTNTTNTT 1907
QY 613 EPTP-----TTPKEAPT-----TPKAAAPTTP--- 635
Db 1908 TTPPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 1967
QY 636 -----KEPAAPTTPKEP-----APTITKSAF-TTPKEAPTTPKEAPTTPKEAPTTP 679
Db 1968 TPTTTTTVTPPT 2027

QY 680 PTTPKKAPKELAPTTTKEP-----TSTTSKAPPT----- 710
Db 2028 PTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2087
QY 711 -----TPKGT-----APTTPKEAP-----TTPKEAPT----- 734
Db 2088 TTTVTPPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2147
QY 735 -----TPKGT-APT-----LKEPAPTTPKAPKELAPTTTKEG 767
Db 2148 TPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2207
QY 768 P-----TSTTSKAP-----TTPKEAPTTPKEAP----- 794
Db 2208 PTPPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2267
QY 795 -----TTPKAPPT-----TPETP-----PTTSVS-----TP 818
Db 2268 PTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2327
QY 819 TTTKEPTTIHKSPESTPELSAETPKALENSPKP-----GVPT-TKTPAATKP 867
Db 2328 TGTQPTT-----TPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2384
QY 868 EMITTAOKDTERDLR-----TTP-----ETTTAAPKMT-----KETATTTKTESKI----- 911
Db 2385 TTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2444
QY 912 -TATTQVSTT-----TQDTPPKITT-----LKTTP 938
Db 2445 TQTPPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2504
QY 939 LAPVTTT-KKTIITTIMKPEETAKPKORATNSKATP-----KPKQPT 983
Db 2505 VTPPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2564
QY 984 KAP-----KKPTSTKKKIMPRVKPTTTPKMTSTMPKMTSRIAEAMLOT 1033
Db 2565 TPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2616
QY 1034 TT-----RPNQTPNSKLVEVNPKSEDAG-----GAE 1059
Db 2617 TTTVTPPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2676
QY 1060 GETPHMLLRHVFMEVTP 1078
Db 2677 TPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2676

RESULT 7

Q9VR49 ID Q9VR49 PRELIMINARY; PRT; 1225 AA.
AC Q9VR49; 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE CG3047-PA.
GN Name=Sgs1; ORFNames=CG3047;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blaziel R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Gabor G.L.,

RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fiertera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.G., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nussekern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang K., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of *Drosophila melanogaster*."
 RT Science 287:2185-2195(2000).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=22426065; PubMed=12537568;
 RA Celisner S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richardson M., Sodergren E.J.,
 RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*
 RT melanogaster euchromatic genome sequence."
 RL Genome Biol. 3:RSESEARCH0079-RESEARCH0079(2002).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celisner S.E.;
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
 RT a genomics perspective."
 RL Genome Biol. 3:RSESEARCH0084-RESEARCH0084(2002).
 RN [4]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=22426069; PubMed=12537572;
 RA Mitra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.J., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celisner S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review."
 RL Genome Biol. 3:RSEARCH0083-RESEARCH0083(2002).
 RN [5]
 RN SEQUENCE FROM N.A.
 RP FLYBASE;
 RG Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RN SEQUENCE FROM N.A.
 RP FLYBASE;

RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB003575; AAF50957.3; --
 DR FlyBase; FBgn0003372; Sgsl;
 SQ SEQUENCE 1225 AA; 127735 MW; A99AF9D4404C78F4 CRC64;
 Query Match 23.0%; Score 1352.9; DB 2; Length 1225;
 Best Local Similarity 27.5%; Pred. No. 12e-18;
 Matches 366; Conservative 121; Mismatches 422; Indels 421; Gaps 53;
 QY 10 LLLLLSFFVQQV-----SSODLSSCAGCGEGYSDATCNDYNCQHYMECCPDF 60
 DB 6 IFUTVSILLIQVKNVKNVNDWSMD-----GFSEEIIPCGGDTIY-----PD- 50
 QY 61 KRYCTAELSKGRCFESFERGECDCDAOCKYDKCCPDYESFCBAEVHNPTSPSSKAP 120
 DB 51 -----PVQPCDIDS-----NPTTKRQKTK 71
 QY 121 PRSGASQTIKSTKSPKPKKKYKVVESIEITEKYKDKNKKKKKPPKPPVVD 179
 DB 72 RP-----KSTRRTTKRT-KRPRKTKTKWK-----RATKRTTKRTRRRRTTPKTP- 117
 QY 180 AGSGLDNGDFKVTTPDT-----STTQHNKVS-----TSPKITTAKPINRPS 221
 DB 118 -----DTTDSPIITGAECTCSDRITASDSDSTTDRTTVNTDWTPLCTDPEPCT- 168
 QY 222 LPNSDTSKETSITVNNKETTVE-----KETTINKQTSDDGKEKT--TSAKTQ 269
 DB 169 -----CSSESSTAIPSSPCIDTSTVTPSCQBTETPTPTCSOGTQTTPCTCAQTT- 221
 QY 270 SIKTSAKDLAPTSKVLAKPTPKAETT--TKGPALTTPKEPTTPKGPASTTPKEPTPT 327
 DB 222 -----TTPSITTTTSRPTTTTPRSITTTTTSRPTTTTTPRSITTTT 262
 QY 328 TKSAPTTKPEAPTTTKSAP--TTPKEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKS 385
 DB 263 TTRRPTTTTPRCITTTTSCAPTTPPTPRSTTTTTSRPTTTTPRCITTTTSCPTTPRS 322
 QY 386 APPT-PKEAPPTP-----KKPAPTTPKPE-----PAPTTPKPE-----P 417
 DB 323 TTTTTSRPTTTTPRCITTTTPSTCTCRPTTTTPRSTTTTCTSGPTTTTPRSTTTTCTSGP 382
 QY 418 TPTTPKEAPT-TKEPAPTTKPEAPTPAPKAPTTPKPE-----PAPTTKPEAPTT 468
 DB 383 TTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTTPRSTTTT 442
 QY 469 TKEPSTTPKPEAPTTTKSAPTTKPEAPTTTKSAP--TTPKEPSTTTKPEAPTTKPE- 525
 DB 443 TSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTTPRST 502
 QY 526 -----PAPTTPKK-----PAPTTPKPE-----PAPTTKPEAPTTTKPAPTA 562
 DB 503 TTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTTPRSTTTTSGPTTTT 562
 QY 563 PKE-----PAPTTKETAPTTPKLTTPTEKLAPTTPKPEAPTTPEELAPTTPEP 614
 DB 563 PRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTTPRSTTTTSGPT 622
 QY 615 TPTTPEE-----PAPTTPKAAAPNTKPEAPTTPKPE-----PAPTTPKPE----- 653
 DB 623 TTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTTPRSTTTT 682
 QY 654 ---PAPTTPKETAP-----TTPKGTAPTTKPEAPTTPKK-----PAPKEL 691
 DB 683 TSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTTPRSTTTTSGPTTTTTPRST 742
 QY 692 APPTTKEPTSTT-----SDKAPPTTPKGTAPTTKPEAPTTPKPE-----PAPTT 735
 DB 743 TTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTTPRSTTTTSGPTTTT 802
 QY 736 PKGTAPTTKPEAPTTPKK-----PAPKELAPTTTKGPTSTT-----SNKP 776
 DB 803 PRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTTPRSTTTTSGPT 862

```

QY 777 APTPKETAPTKPAPTPPKK-----PAPTPPET-----PPPTTSEVSTPTT 820
DB 863 TTTTTPSTTTTCTSGPTTTPRSTTTTCTSCPTTTPRSTTTTCTSCPTTTPRSTTTTCT 922
QY 821 TKSPETTHKSPDESPELSAETPKALENSPKPGVPTKPAAPKPEMTTAKDKTTER 880
DB 923 TSGPTT--TTPRSTTKTSCAPT-----TTPRSTTTTTSRPTTTPRSTTTT-----TTSR 973
QY 881 DURTTPETTTAAKMKETATTEKTTESKITATTQVTSITQDTTTPFKITLTKTTTLLA 940
DB 974 PTTTTPRSTT--TPSTRPTTTPRSTT-----TTSRPTTTPRSTT-----KTSICA 1021
QY 941 PKVTTTKTITTEIMNKPEETAKPKDRATNSKATPKPKQTKAPK-----KPTST 992
DB 1022 PTTTTPRSTTTT-----TSRPTTTPRSTTTTTSRPTTTPRSTTTPCPTSRTT 1073
QY 993 KKPWKMPVRVKETTPTRKMT-----STWPELNPTSRIAEMLOTTTR-----1036
DB 1074 TPRSITTTTTSRPTTTPRSTTTPCPTTPSPR-----TTPTRPCPCHPQPYQ 1126
QY 1037 -----PNOT-----PNSKLVEVNPKSEDAGGAGETPHMLRPHVFMPEVTPD 1079
DB 1127 IPPWSWVNPPTYPNVPWFQPN-----PVPVQ 1154
QY 1080 MDYLPVRPNQ 1089
DB 1155 WPMQGYPNQ 1164

RESULT 8
QY Q7KTF6 PRELIMINARY; PRT; 1761 AA.
AC Q7KTF6
DC Q7KTF6
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DE CG33300-PA (Fragment).
GN ORFNames=CG33300;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beson D.Y., Betsun P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matti B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle B.J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

```

```

RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Javerty T., Muzny D.M., Nelson C.R.,
RA Pacle B.J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RA "Finishing a whole-genome shotgun: release 3 of the Drosophila
RL melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RA "The transposable elements of the Drosophila melanogaster euchromatin:
RL a genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochownik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RA "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RX FLYBASE;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX FLYBASE;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB003626; AAS64673.1; -
DR InterPro; IPR011009; Kinase_like.
FT NON TER 1
SQ SEQUENCE 1761 AA; 197521 MW; 62A5E16B6241B3F4 CRC64;

```

```

Query Match 22.5%; Score 1327.5; DB 2; Length 1761;
Best Local Similarity 24.7%; Pred. No. 1.le-17;
Matches 451; Conservative 114; Mismatches 369; Indels 895; Gaps 85;
QY 74 CF-----ESFGRGECDCDAQCKYDKCPDYEGSFC---104
DB 5 CFIWALYHSANNALEGSBIKFGDSESTISAE-----KY-----PY-----CWIN 45
QY 105 -----AVHNPTSPPS 115
DB 46 PELPGCPFGPGDGNNTTKISTLATTKSTVTTSEETTLKTIKSTAKPTTKTNEPT 105
QY 116 SKKAPPPSCA-----SQTIKST--TKR 135
DB 106 TEKTIPTKATTKTATTATATATAPKTEQTTLRTTIKSTSELITLTKTTIKSTAEATKK 165

```

QY 136 SPKPPNKKTKVIESEBITEKVDKNKKNKTKKPTKPPVDEAGSLDNGDFKVT-- 193
 DB 166 STNPTTKKSTLRTTEPTTRKSSTAK--TTREPTTK--RETTTER 207
 QY 194 ---PDTS--TTCNKNVSTSPKITTAKINPRPSLPNSDTSKTSLVNKETVE-- 243
 DB 208 TQEPSTSKTTHE---TAEPAATKTTHEPT--TQSKTLRITEPTTRKSSTA 257
 QY 244 --TKETTTNQTSDGKEKTSIAKQSIKTSKADLAPTSKVLAKPTPKAETTTKGA 301
 DB 258 KTTREPTTKRTTERTKEPST---KTTTBE--TAEPAATKTTHEPTTQKSTI 307
 QY 302 LTTTKEPT-----PTTPKEPATTPKPTPTTKTSAPTTKPEA-----PT 342
 DB 308 LRITEPTTRKSSTARTTRETTRTERTTQEPSTKTHETT--ABPATKTTHEPT 366
 QY 343 TTGSA-----PTTPKSA-----PTTPKEPATTPPKPA----- 379
 DB 367 TQKSTLIRITEPTTRKSSTAKTTRE--PTTKRTTERTTKEPSTKTTTHTAABATK 424
 QY 380 -----PTTPKSA-----PTTPKEPATTPPKPA----- 402
 DB 425 KTTHEPTTQKSTLIRITEPTTRKSSTAKTTRETTRTERTTQEPSTKTHETTA 484
 QY 403 -----PTTPKSA-----PTTPKEPATTPPKPA-----PTT 436
 DB 485 EPATKTTHEPTTQKSTLIRITEPTTRKSSTAKTTRETTRTERTTQEPSTKTTT 544
 QY 437 PKEPATAPKPA--PTTPK-----EPA-----PTTPKEPATTPPKPS-- 473
 DB 545 HETTAETATKTTHEPTTQKSTLIRITEPTTRKSSTAKTTRETTRTERTTQEPST 604
 QY 474 -----PTTPKSA-----PTTPKSA-----PT-- 490
 DB 605 SKTTTHTAEPATKTTHEPTTQKSTLIRITEPTTRKSSTAKTTRETTRTERTT 664
 QY 491 -----TTKPEA-----PTTPKSA-----PTTPKEPATTPPKPS-- 510
 DB 665 QEPSTSKTTTHTAEPATKTTHEPTTQKSTLIRITEPTTRKSSTAKTTRETTRT 724
 QY 511 SPTTPKPEA-----PTTPK-----EPA-----PT 529
 DB 725 TERTTKEPTTRKTTTHTKTTTEPTTKTTTKEPTTKKSTLTKPTEPTTRKTTTKEPT 784
 QY 530 TPKEPATTPKPEA-----PTTPKEPATTPPKPA-----PTAPK-----EPA 567
 DB 785 TTKRTTERTTKEPTTRKTTTHTKTTTEPTTK--TTTKTTHEPTTKKSTLTKPTEPT 840
 QY 568 -----PTTPKETA-----PTTPKKT-----PTTPKKT----- 583
 DB 841 TRKTTTKEPTTRKTTTKEPTTRTERTTQEPSTKTTTHTAEPATKTTHEPTTQKSTLRI 900
 QY 584 ---PTT--PEKLAPTEKPAPTTPELAPTTPEPT-----PTTPE-----EPA-- 623
 DB 901 TEEPTTRKSSTAKTTRE--PTTKRETTERTKPTTRKTTTHTKTTTEPTTKTTHEPT 957
 QY 624 -----PTTPKAAAPTTPKEPATTPPKPEPATTPKPEA-----PTTPK 660
 DB 958 KKSSTLTKPTEPTTRKTTSTTTTRE--PTTKRTTERTTKEPTTRKTTTHTKTTTEPTATNK 1015
 QY 661 ETA-----PTTPKETA-----PTTPKEPATTPPKPAKELA-----PTTKEPTS 701
 DB 1016 TTKTTTHTHEPTTKKSTLTKPTEPTTRKTTSTTTTRETTRTERTTRETPTTKTTTH 1075
 QY 702 TTSKPA-----PTTPKGA-----PTTPKEPATTPPKPEA-----PTTP 736
 DB 1076 ETAEPAKTTTHTTQKSTLIRITEPTTRKSSTAKTTRETTRTERTTKEPTTR 1135
 QY 737 KGTAPTTKPEA-----PTTPK-----KPAKELAPTTTKGPTSTSDKPAPTPKETA 785
 DB 1136 KTTTHTKTTTEPTTKTTTHTPTTKKSTLTKPTEE---PTTRK--TSTTKTRETPTTKRTT 1190

QY 786 PTTKPEA-----PTTPK-----XP-----APTTPE 807
 DB 1191 ERTTKEPTTRKTTTHTKTTTEPTTKTTTHTKTTTHTKTTTHTKTTTHTKTTTHTKTT 1250
 QY 808 PPTTSEVSTPTTTKEP-----TTIHKSPDE-----STPELSAEPKALENSKEPQVP 857
 DB 1251 REPTTKRVTTERTTRETPTTRKTTTHTKTTTEPTTKTTTHTKTTTHTKTTTHTKTTTHTKTT 1309
 QY 858 TTKTAAAKPEMTTAKOTTERDLR-----TTPPE 887
 DB 1310 TTRKSTTTTRETPTTKRVTTERTTRETPTTRKTTTHTKTTTEPTTKTTTHTKTTTHTKTT 1369
 QY 888 TTTAAP-----KTKETAT-----TTEKTESKITATTTQVSTTTQDTPPKIT 932
 DB 1370 STTLKTEPTTRKTTTHTKTTTRETPTTKRVTTERTTRETPTTRKTT--THTKTTTEPTTKTT 1427
 QY 933 TTKT-----TTLAP-----KVTTK-----KTTTTEIMNKP-----EE 961
 DB 1428 TTKTTTHTPTTKKSTLTKPTEPTTRKTTTHTKTTTRETPTTKRVTTERTTRETPTTRKTTTHTKTT 1487
 QY 962 TAKPKDRATNSKATPKQ-----XP-----TKAPKKT----- 990
 DB 1488 TEEPTTKTTTHTKTTTHTKTTTRETPTTKSTLTKPTEPTTRKTTTHTKTTTRETPTTKRVTTERTTRET 1547
 QY 991 -----STKKPKTMPVRKPKTPTPRKMTSTMPELNP 1022
 DB 1548 TTRKTTTHTKTTTHTKTTTHTKTTTRETPTTKKSTL-----KPTTEPTTRKTTTHTKTTTRETPT 1603
 QY 1023 TSRIAEMLQTT-----TRPNOTNSKLAVEVNPKSEDAGGAGET----- 1062
 DB 1604 TTRTSTSVKTTADOTTTRKTTAEMSTTQEPSTVETTTNSQSNSTTESITTEBQVHH 1663
 QY 1063 -----PHMLLRPH----- 1070
 DB 1664 HHHHHYHKPADLGSILPL 1723
 QY 1071 -----VFMPEVT--PDMDYLPRVN 1088
 DB 1724 PPLPEVNTALISLPEISLPL 1752
 RESULT 9
 AAS64673 PRELIMINARY; PRT; 1761 AA.
 ID AAS64673; DT 01-APR-2004 (TREMREL. 27, Created)
 AC AAS64673; DT 01-APR-2004 (TREMREL. 27, Last sequence update)
 DT 01-APR-2004 (TREMREL. 27, Last annotation update)
 DE CG33300-PA (Fragment).
 GN CG33300.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 CX NCBI_TaxID=7227;
 RN [1]
 SEQUENCE FROM N.A.
 RP MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blasej R.G., Champ E.M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Agayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Bouchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Gaig N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hosain D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kianes I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodagef, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT melanogaster euchromatic genome sequence";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomic perspective";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBSJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AE003626; AAS64673.1; -;
FT NON TER 1 1
SQ SEQUENCE 1761 AA; 197521 MW; 62A5E16E6241B3F4 CRC64;
Query Match 22.5%; Score 1327.5; DB 2; Length 1761;
Best Local Similarity 24.7%; Pred No. 1.1e-17;
Matches 451; Conservative 114; Mismatches 369; Indels 895; Gaps 85;

Qy	74	CF	-----ESFERGECDCDAOCKYDKCCPDYESFC	104
Db	5	CFTLALYHSANAALGSEIKFGDESTESA	-----KY-----PDY-----CWIN	45
Qy	105		-----AEVHNPTSPPS	115
Db	46	PFLPGCPGPGDGNSTTKISTLATTKSVTTSBETTLKIITIKSTAKPTTQKTNEPT	105	
Qy	116	SKKAPPPSGA	-----SOTIKST--TKR	135
Db	106	TEKITTPKATIKSTATTARATEAPKTEQTLTKTIKSTSELITLTKTIKSTAEATKK	165	
Qy	136	SPKPPNKKTKKIVIESEITEIKVDKNKNTKKKTPKPPVVDVDEAGSLDNGDFKVT	193	
Db	166	STHNPTKSTLIRITEPTTRKSTAK--TTREPTTK	207	
Qy	194	-----PDTS--TQHNKVSFKITITAKPINRPSLPNSDTSKETSULTVNKEITVE	243	
Db	208	TTQEPSTSKTTH	-----TQSTTLIRITEEPTTRKSSTA	257
Qy	244	--TKETITINKQISTDGEKKTTSKAKTQSIKTSKADLAPTSKVLAKPTKAEATTTKGA	301	
Db	258	KTIREPTTKRETTERTTKPSTS	-----KITTHE--TTAEPATKKTTHTEPTTQKSTT	307
Qy	302	LTPKPT	-----PTTPKEPASTTPKEPTPTTIKSAPTTPKEPA	342
Db	308	LRITEPTTRKSSTARTTRETTRKRETTERTTQEPSTSKTKTHETT--AEPATKKTTHET	366	
Qy	343	TTKSA	-----PTTPKEPAPTTPKEPAPTTPKEPA	379
Db	367	TQKSTLIRITEPTTRKSTAKTIRE--PTTKRETTERTTKPSTS	424	
Qy	380	-----PTTKSA	-----PTTPKEPAPTTPKKA	402
Db	425	KTTHTEPTQKSTTLIRITEPTTRKSTARTTRETTRKRETTERTTQEPSTSKTKTHETTA	484	
Qy	403	-----PTTPKEPA	-----PTTPKEPTTPKEPA	436
Db	485	EPATKKTTHTEPTQKSTTLIRITEPTTRKSTAKTTRBPTTKRETTERTTQEPSTSKTTT	544	
Qy	437	PKEPAPTAPKKA--PTTPK	-----EPA	473
Db	545	HETTAETATKKTTHTEPTQKSTTLIRITEPTTRKSTAKTTRBPTTKRETTERTTQEPST	604	
Qy	474	-----PTTPKEPA	-----PTTKSA	490
Db	605	SKTTHTETTAEPATKKTTHTEPTQKSTTLIRITEEPTTRKSTAKTTRBPTTKRETTERT	664	
Qy	491	-----TTKEPA	-----PTTKSA	510
Db	665	QEPSTSKTTHETTAEPATKKTTHBPTTQKSTTLIRITEPTTRKSTAKTTRBPTTKRET	724	
Qy	511	SPTTKEPA	-----PTTPK	529
Db	725	TERTTKETTRKKTTHKTTBPTTKKTTHTEPTTKKSTTLKPTBPTTRKSTTKTTRBPT	784	
Qy	530	TPKKPAPTTPKEPA	-----PTTPKEPAPTTPKKA	567
Db	785	TKRKTERTTKETTRKKTTHKTTBPTTK	-----TTTKKTTHTEPTTKKSTTLKPTBPT	840
Qy	568	-----PTTPKEPA	-----PTTPKKLT	583
Db	841	TRKSTTKTTRBPTTKRETTERTTQEPSTSKTTHETTAEPATKKTTHTEPTTQKSTTLRI	900	
Qy	584	---PTT--PEKLAPTPEKAPTTPPELAPTPEPT	-----PTTPE	623
Db	901	TEPTTRKSTAKTIRE--PTTKRETTERTTKETTRKKTTHKTTBPTTKKTTHTEPTT	957	
Qy	624	-----PTTPKAAAPTTPKEPAPTTPKEPA	-----PTTPK	660
Db	958	KKSTTLKPTBPTTRKSTTKTIRE--PTTKKKTERTTKETTRKKTTHKTTBPTTK	1015	
Qy	661	ETA	-----PTTLKEPAPTTPKAPKELA	701

```

DB 1016 TTGKTTTHTTTGKSTLTKPTEPTTRKSTTKTTRTEPTTKRVTTERTREPTTSKTTTH 1075
QY 702 TTSDKFA-----PTTPKGA-----PTTKPAPPTPKPA-----PTTP 736
DB 1076 ETTAEPATKTTHTTTGKSTLTKPTEPTTRKSTAKTTRTEPTTKRTEPTTKTTR 1135
QY 737 KGTAPTTKPA-----PTTPK-----KPAKELAPTTTKGPTSTTSDDKAPPTPKETA 785
DB 1136 KTTHTKTEPTTKTTHTEPTTKSTLTKPTEE-----PTTRK-----TSTTKTREPTTKKTT 1190
QY 786 PTPKPA-----PTTPK-----KP-----APTTPET 807
DB 1191 ERTTEPTTRKTTTHKTEPTTKTTHTEPTTKSTLTKPTEPTTRKSTTKTT 1250
QY 808 PPTTSVSTPTTKP-----TTTKSPDE-----STPELSABPTPKALENSKPEGVP 857
DB 1251 REPTTKRVTTERTREPTTRKTTTHKTEPTTKTTHTEPTTKTTHTEPTTKK-STILKPEEP 1309
QY 858 TTTPAANKPSTTTAKDTHTERDLR-----TTPE 887
DB 1310 TTKTSTTKTREPTTKRVTTERTREPTTRKTTTHKTEPTTKTTHTEPTTKKTTTHTEPTTK 1369
QY 888 TTTAAP-----KMTKETAT-----TTKETESKITAITTQVTSITTTQDTPPKIT 932
DB 1370 STILKPEPTTRKSTTKTTRTEPTTKRVTTERTREPTTRKTT-THKTEPTTKTT 1427
QY 933 TLKT-----TTLAP-----KVTTK-----KTITTEIMNKP-----EE 961
DB 1428 TKKTTHEPTTKSTLTKPTEPTTRKSTTKTTRTEPTTKRVTTERTREPTTRKTTTHKT 1487
QY 962 TAKPKDRATNKATPKPQ-----KP-----TKAPKKPT-----990
DB 1488 TEEPTTKTTTKTTHTEPTTKSTLTKPTEPTTRKSTTKTTRTEPTTKRVTTERTREPT 1547
QY 991-----STKKPKTPRVKPKTTPFRKMTSTMPELNP 1022
DB 1548 TTRKTTTAAHTTEPTTKTTHTEPTTKSTLTKPTEPTTRKSTTKTTRTEPTTKSTTKTTR 1603
QY 1023 TSRIAEAMLOTT-----TRNQTPNSKLAVNPKSEADAGAGET-----1062
DB 1604 TTRETSTVKTTADQTTKRTAEMSTTNQEPSTVETTTNNSNQNTTETSTTBEQVHH 1663
QY 1063-----PHMLLRPH-----1070
DB 1664 HHHHHYHKPADLGSILPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPL 1723
QY 1071-----VFMPREV-PMMDYLPRVN 1088
DB 1724 PPLPEVNLTALSLPEISLNPPLPQLPN 1752

```

RESULT 10

Q7PMD5

ID Q7PMD5 PRELIMINARY; PRT; 3150 AA.

AC Q7PMD5; 01-MAR-2004 (TrEMBLrel. 26, Created)

DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE ENSANGP0000004655 (Fragment)

DE Name=ENSANGG0000003651;

OS Anopheles gambiae str. PEST

OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.

OX NCBI_TaxID=180454;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PEST;

RA Anopheles Genome Sequencing Consortium;

RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.

CC !- CAUTION: The sequence shown here is derived from an

CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is

CC preliminary data.

```

DR ENBL; AAB0100980; EAA13969.2; --
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0008151; P:cell growth and/or maintenance; IEA.
DR InterPro; IPR006770; OGF recept.
DR Pfam; PF04680; OGF_r_III; 80.
FT NON_TER 1
FT NON_TER 3150 3150
SQ SEQUENCE 3150 AA; 322879 MW; 3C7B3D441CB5C839 CRC64;

Query Match 21.5%; Score 1266.3; DB 2; Length 3150;
Best Local Similarity 20.1%; Pred. No. 3.2e-16;
Matches 411; Conservative 110; Mismatches 416; Indels 1107; Gaps 76;

QY 110 PTPSP-----SSKAPPPSGASQIKSTKSP-KPPNKKTKKVISEETEKVK 159
DB 832 PTPSTSTSTDTMTSSASTPEPS-----TPGTRTTRTPTSTSTDTMTSSASTPEST 887
QY 160 DNKKNR-TKKKPTKPPVVDKAGS-----GLDNGDFKVTTPDTS 197
DB 888 TSGTTRTTRTPTDITMTSSASTPESTTPGTRTTRTPTSTSTDTMTSSASTPEPS 947
QY 198 TTQHN-----KVTSPKITAKPINP-----218
DB 948 TTPGTRTTRTPTSTSTDTMTSSASTPESTTPGTRTTRTPTSTSTDTMTSSAST 1007
QY 219-----RPSLPNDSKTSKSLVNVKETTVEKETTNNKQSTDKGKKT 263
DB 1008 PPSKPGTRTTRTPTSTSTDTMTSSASTPESTTPGTRTTRTPTSTSTDTMT 1066
QY 264 SA-----KETQSIEKTSKDLAP-----TSKVLAKPPEKPT 295
DB 1067 SSASTPESTTPGTRTTRTPTSTSTDTMTSSASTPESTTPGTRTTRTPTSTST 1126
QY 296-----TTKGPAITPKETP-----TTPEKAS 318
DB 1127 DITMTSSASTPESTTPGTRTTRTPTSTSTDTMTSSASTPESTTPGTRTTRTPT 1186
QY 319-----TPKPEP-----PTTKKSAPT 334
DB 1187 TSTSTDTMTSSASTPESTTPGTRTTRTPTSTSTDTMTSSASTPESTTPGTRTTR 1246
QY 335 TP-----KEPA-PTTKKSAPT-----351
DB 1247 RPTSTDTMTSSASTPESTTPGTRTTRTPTSTSTDTMTSSASTPESTTPGTRTTR 1306
QY 352-----KEPAPTTPKPAPTTPKPAPT-----TTKEPA-PTTT 383
DB 1307 TRPTSTSTDTMTSSASTPESTTPGTRTTRTPTSTSTDTMTSSASTPESTTPGTR 1366
QY 384 KSAPTTP-----KEPAP-----TPKPKAPT-----TPKEPAPTTP 414
DB 1367 RTTPTRPTSTSTDTMTSSASTPESTTPGTRTTRTPTSTSTDTMTSSASTPEPTT 1426
QY 415 KEPTTPTPKPAPT-----TKEPAP-----TPKPAPT-----APKKA 449
DB 1427 PGTRTTRTPTSTSTDTMTSSASTPESTTPGTRTTRTPTSTSTDTMTSSASTPE 1486
QY 450 P-----TPKPEP-----TPKPAPTTPKPSPTTP 477
DB 1487 PSTTPGTRTTRTPTSTSTDTMTSSASTPESTTPGTRTTRTPTSTSTDTMTSSAST 1546
QY 478 KEPA-----PTTKKSAPT-----TKEPAP-PTTKKSAPT-----507
DB 1547 -PESMTPGTRTTRTPTSTSTDTMTSSASTPESTTPGTRTTRTPTSTSTDTMTS 1605
QY 508 --KEPSPTTPKPAPTTPKPAPT-----TPKKA 536
DB 1606 SASTPESTTPGTRTTRTPTSTSTDTMTSSASTPESTTPGTRTTRTPTSTSTDTMT 1665
QY 537-----TPKPAPT-----TPKEPAPTTPKKA-----PTA 562
DB 1666 SSASTPESTTPGTRTTRTPTSTSTDTMTSSASTPESTTPGTRTTRTPTSTSTDTMT 1725

```

QY 563 P-----KEPATTTPKETAATTPKLTTP-----TPE-----KL 590
Db 1726 PTSTESTDTMTSSASTPEPSTTPGTRTTPRPTSTESTDTMTSSASTPEPSTTPGTRT 1785
QY 591 APT-----TPE-----KPAATPEE 605
Db 1786 TTPRPTSTESTDTMTSSASTPEPSTTPGTRTTPRPTSTESTDTMTSSASTPEPSTTPG 1845
QY 606 LAPTPPEPTT-----TPE-----EPAPT 625
Db 1846 TTPRPTSTESTDTMTSSASTPEPSTTPGTRTTPRPTSTESTDTMTSSASTPEPST 1905
QY 626 TPKAAAPNTPKEPAP-----TPKEPAP----- 648
Db 1906 TTPGTRTTPRPTSTESTDTMTSSASTPEPSTTPGTRTTPRPTSTESTDTMTSSASTPEP 1965
QY 649 -----TPKEPAPTTPKETA-----PTPKGTAPATTLKAPAP----- 680
Db 1966 SMTPTGTRTTPRPTSTESTDTMTSSASTPEPSTTPGTRTTPRPTSTESTDTMTSSAST 2025
QY 681 -----TPPKKAPKELAPTT-----TKPTST-----TS 704
Db 2026 TPEPSTTPGTRTTPRPTSTESTDTMTSSASTPEPSTTPGTRTTPRPTSTESTDTMTSSA 2085
QY 705 DKPAPTTPKGTAPTPKEP-----APTP----- 728
Db 2086 STPEPSTTPGTRTTPRPTSTESTDTMTSSASTPEPSTTPGTRTTPRPTSTESTDTMTSS 2145
QY 729 -----KEPAPTTPKGTAPATTLKAPAP-----TPPKKAPKELA-- 761
Db 2146 SMSSASTPEPSTTPGTRTTPRPTSTESTDTMTSSASTPEPSTTPGTRTTPRPTSTESTDT 2205
QY 762 -----PTTKGTSTTSOKPAPT-----TPKE-----TAPTTPKEPAP- 794
Db 2206 TMSSASTPEPSTTPGTRTTPRPTSTESTDTMTSSASTPEPSTTPGTRTTPRPTSTESTDT 2265
QY 795 -----TPPKKAPAP-----TPPE----- 806
Db 2266 DTTMTSSMSSASTPEPSTTPGTRTTPRPTSTESTDTMTSSMSSASTPEPSTTPELLQDD 2325
QY 807 ----- 806
Db 2326 HLPPOCRLVCROPLRSHRHLVQPELLQDDHLPPOCROPLRSHRHLVQPELLQD 2385
QY 807 -----TPPPTT-----SEVSTPTTKPTTHKSP- 831
Db 2386 DQHLPTSTDTMTSSASTPEPSTTPGTRTTPRPTSTESTDTMTSSASTPEPSTTPGTRTTP 2445
QY 832 -----DESTPELSAP-----TPKALE 848
Db 2446 RPTSTESTDTMTSSASTPEPSTTPGTRTTPRPTSTESTDTMTSSASTPEPSTTPGTRT 2505
QY 849 NSPKEP-----GVPTTKTPAATPEMTTAKDTERDLRTTPETTTAAKMTKETATT-- 902
Db 2506 TTPRPTSTESTDTMTTAST-PEPSTT-PDTRTTPRPTPTDPTMTSSASTPEPSTTPG 2563
QY 903 -----TEKTESKITATTQVTSITTTQDTTPFKIT 933
Db 2564 TTRTTPRPTPTDPTMTSSASTPEPSTTPGTRTTPRPTSTESTDTMTSSMSSASTPEPSTT 2623
QY 934 LKTTTLAP-----KVTTTKITTTTEIMNKPEETAK 964
Db 2624 PGTRTTPRPTPTDPTMTSSMSSASATPGTRTTPRPTSTESTDTMTSSASTPEPSTT 2683
QY 965 PKDRATN-----SKATTPKQ-----KPT----- 983
Db 2684 PGTRTTPRPTSTDTMTSSMSSASTPEPSTTPGTRTTPRPTPTDPTMTSSMSSAST 2743
QY 984 -----KAPKPTSTK-----KPKMP-----RVRKPKTTPPKWTSTM----- 1017
Db 2744 TPGTRTTPRPTSTESTDTMTSSASTPEPSTTPGTRTTPRPTPTDPTMTSSMSSAST 2803

QY 1018 --PELNP-TSRIAEAMLOTTTRNQTPNSKLVENPKSESDAGGAEGTTPHMLLRPHVMP 1074
Db 2804 PEPSTTPGTRT-----TTPRPTPTDSTMTSSMSAS-----TPE-----P 2839
QY 1075 EVTP 1078
Db 2840 STTP 2843
RESULT 11
Q6SSE6 PRELIMINARY; PRT; 3409 AA.
AC Q6SSE6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Plus agglutinin.
GN Name=SAG1;
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
RA Goodenough U.W.;
RL Submitted (Oct-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY450704.1; -;
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR003882; Pistil_extensin.
DR PRINTS; PR01222; ATROPHIN.
DR PRINTS; PR01218; PSTLEXTENSIN.
SQ SEQUENCE 3409 AA; 336045 MW; 7FE87633EDD6631F CRC64;
Query Match 20.5%; Score 1205.1; DB 2; Length 3409;
Best Local Similarity 15.2%; Pred. No. 5.2e-15;
Matches 347; Conservative 143; Mismatches 400; Indels 1399; Gaps 61;
QY 1 MAWTLPT-----YLLLLLV----- 16
Db 50 LAWVTIPVSDALQPPVNVSVTSSAATDAPSAIAYIGALLDVLNVPQWQNCINDRRSYD 109
QY 17 ----- 16
Db 110 APWFSRCPALPAVQGYDEYDITVYSGNSLRPFYSTCRYPASSDPYAFLEPYSMFWD 169
QY 17 -----FVIQVSS-----QDLSSCAGRCGEGYSRD 41
Db 170 DFGIYQPVLDGFDMMWGFTVNTVDSNLKWDIPAWTAQGAWLGGQDF-----RD 217
QY 42 ATCN-----CDYNCOHYMECPDFKRVCTAELSCKRCFESFERGECDDAOCKYD-- 94
Db 218 AVWNFGTHYCSW----PVEEC-----SSCENIDIA 243
QY 95 ----- 94
Db 244 DPYDPDKIANGIVPAVITALDFRNASLYIYVLDGFMFAGSLDNVWYLNLAYNFIGGFLP 303
QY 95 -----KCCPD-----YE----- 101
Db 304 ANLPTLLPSLOHALDHCRAFPDVRGTASLOYGWQVYPTGQVYECSDGVTGSDGTEY 363
QY 102 ----- 101
Db 364 VISGMIPEWDGDAVAAAASDSSLPWANLRTVRLSNQALYGPILPGLRSASSISSWRLQG 423
QY 102 -----SFCAEVHN----- 109
Db 424 NTELCGPLPEFAAPINSLLYLGLTHTVWHDPAHQEGECIAPPSPPPSPRPPR 483
QY 110 -----PTSPSSKKAP-----PPSGASQT----- 128
Db 484 PPPLPPSPPPPLLPSPDPVPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 543

[illegible]

QY	874	-----KDKTTERDL-----RTPP-----	886
Db	1513	WGPGTETAABWADVDGTYQLQLSIGGVYTRTVVVDSTPPSVSGNVTLSANRIKQEP	1572
QY	887	-----ETTTAAPKMT-----	896
Db	1573	SAYGEASLNALGSKQAMLLTISEPVPAPFDPAASLIIVTCALVAEWAADKMTFFVILAM	1632
QY	897	---KETAATTEKTESKIT---	912
Db	1633	TLPAELVATAAGSSSSSGTSRSGNGNGTAAAAAAPPAGTGTGRRRALQOQAAAAPPPP	1692
QY	913	-----TTTKTITTE-----	912
Db	1693	ASGSSSLGAAATANCOORHVHFLPATAYADAARNPGRNDLSLSVELTDNAVASPAVGE	1752
QY	913	--ATTQVSTTTQDTPPKITLTKTTTLAPKV---	944
Db	1753	ALATTARVTAATYP-----AVAATTTLVAAAASSFAQAIRAKGSLLOGSYHIQMLTWS	1805
QY	945	-----TTTKTITTE-----	954
Db	1806	LYLASRGVGREYGEYAVEFKAVLGVKGNLGPAAAMPITNEKEVTAABQARQVGGDLWPI	1865
QY	955	---IMNKPBEATAKDRATNSKATTPKPKP---	982
Db	1866	GNLGGSSNTTASGSSSSSSSSNSP-PRRPPPPPAAGSTGLLFSNADASPPPLAVATP	1924
QY	983	-----TKAPKPTSTKPKTWPRVKPTTPTPKMTST--MPELN-----	1021
Db	1925	ALPAPLPSTIAAATAAP-----PRLPSP--PPPVGSGTGVLPRHLMQOM	1969
QY	1022	-----PTGRIABAMLQTTTRPNQTPNSKIVE-----	1047
Db	1970	LQPPAAAAVAPP PPPP ASSSALVLQPPPPPPPPSLLIQOASATYVSDMQDLLYTLVWAA	2029
QY	1048	-----VNPKSEDAGAGETPHMLLRHVHVPPEVTPDMVDYLP--	1085
Db	2030	MLTVAAGRLIAAVLYRLLVSP-----PHPPFLA-----	2064
QY	1086	-----VP	1087
Db	2065	TTIAGLILVALTFYSCWALGGPAADWHGSRTAAYCVLTIAVVPYAAFLMWLALARAMMVP	2124
QY	1088	NOGIINPM 1096	
Db	2125	-QFTLVEPM 2132	
RESULT 12			
AAS07044			
ID	AAS07044	PRELIMINARY;	PRT; 3409 AA.
AC	AAS07044;		
DT	02-MAR-2004	(T-EMBLrel. 27, Created)	
DT	02-MAR-2004	(T-EMBLrel. 27, Last sequence update)	
DT	02-MAR-2004	(T-EMBLrel. 27, Last annotation update)	
DE		plus agglutinin.	
GN		SAG1.	
OS		Chlamydomonas reinhardtii.	
OC		Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;	
OC		Chlamydomonadaceae; Chlamydomonas.	
CC		NCBI_taxid=3055;	
CX		[1]	
RN		SEQUENCE FROM N.A.	
RP		Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,	
RA		Goodenough U.W.;	
RT		"Plus and Minus Sexual Agglutinins from Chlamydomonas reinhardtii.;"	
RL		Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.	
DR		EMBL; AY450930; AAS07044.1; --	
SQ		SEQUENCE 3409 AA; 336045 MW; 7FE87633EDD6631F CRC64;	
Query Match 20.5%; Score 1205.1; DB 2; Length 3409;			

QY 1048 -----VNPKESEDAGAGETPHMLLRPHVFMVPTDMDYLPR-- 1085
 Db 2030 MLITAVAGRLTAANVRLVSP-----PFPFLA-----FPRLE 2064
 QY 1086 -----VP 1087
 Db 2065 TTIAGLILVALTFYSCMALGSPADWHSRTAAVCVLITAVVYAAFLWLALARAMVVP 2124
 QY 1088 NOGLIINPM 1096
 Db 2125 -OFTLVEPM 2132

RESULT 13
 SLPI_CLOTM
 ID SLPI_CLOTM STANDARD; PRT; 1664 AA.
 AC Q06852;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer protein 1).
 DE protein 1).
 GN Name=olpB;
 OS Clostridium thermocellum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OC NCBI_TaxID=1515;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIB 10682;
 RX MEDLINE=93209931; PubMed=8458832;
 RA Fujino T., Beguin P., Aubert J.-P.;
 RT "Organization of a Clostridium thermocellum gene cluster encoding the cellulosomal scaffolding protein C1PA and a protein possibly involved in attachment of the cellulosome to the cell surface."
 RL J. Bacteriol. 175:1891-1899(1993).
 CC -!- SUBUNIT: Assembled into mono-layered crystalline arrays.
 CC -!- SUBCELLULAR LOCATION: Cell wall.
 CC -!- SIMILARITY: Contains 4 S-layer homology (SLH) domains.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X67506; CAA47841.1; -.
 DR PIR; T18262; T18262.
 DR InterPro; IPR008965; Cellul_bind.
 DR InterPro; IPR001119; SLH.
 DR Pfam; PF00395; SLH; 3.
 DR PROSITE; PS01072; SLH DOMAIN; 2.
 DR Cell wall; Repeat; S-layer; Signal.
 KW SIGNAL 1 28
 FT CHAIN 29 1664
 FT DOMAIN 36 763
 FT REPEAT 36 191
 FT REPEAT 207 363
 FT REPEAT 409 565
 FT REPEAT 607 763
 FT DOMAIN 771 1377
 FT DOMAIN 1378 1449
 FT DOMAIN 1453 1494
 FT DOMAIN 1495 1565
 FT DOMAIN 1566 1625
 FT DOMAIN 1626 1646
 SQ SEQUENCE 1664 AA; 178194 MW; 5F396695BA9FE74B CRC64;

Query Match
 Best Local Similarity 19.2%; Score 1132; DB 1; Length 1664;
 Pred. No. 5.3e-14; Pred. 1664;

Matches 371; Conservative 124; Mismatches 421; Indels 820; Gaps 69;
 QY 4 KTLPIYLLILLVVFVQVSSQDLSSCAGCGEGYSDATCNCYNCQHYMECCP----- 58
 Db 6 KVLISILUTLL-----IISTTSVNM-----FAEATPSIEW 37
 QY 59 -----DFKEVCTAELSKRCFCPEFSGRECDCAQCKYDKCCPD 99
 Db 38 LDKTEVHVGDVITATIKVNNIRKLAGYQLNIK----- 69
 QY 100 YESFCAEVHNPTSP-----PSSKAPPPSGASQTIKS----- 131
 Db 70 ---FDEVLQFVDPATGEEFTDKSMFNVRLTNSXYGTPV-AGNDIKSGINPATGYN 125
 QY 132 --TKGRSP-----KPPNKKTK----- 146
 Db 126 NLTAYSKGIDEHTGIIGIFKVLKQNTSRFEDTSLMPGAISGTSLPDWAETITGY 185
 QY 147 -----KVIESEITE-----KVK-----DNKK-----RTKKKPT- 171
 Db 186 EVIQPOLIVVEABPLKDAVLELDKTKVKVGDIITATIKIENMKNFAGYQLNIKVDPTM 245
 QY 172 -----PKPPVVDGAGSLDN----- 186
 Db 246 LEAIELETGSAIAKRTWPTVGTGVLQSDNYGKTTAVANDVGAGIINFABAYNLTKYRT 305
 QY 187 -----GDFKVITPDT-----STTQHNKV 204
 Db 306 GVAETGIIKIGIFRVLKAGSTAIRFEDTTAMGAIEGTWFDWYGENIKGVSVQPGI 365
 QY 205 -----SISPKITAKPINRPSLP-----PNS-----DTSK----- 230
 Db 366 VAEGEPGEEPTPEPTVDTPTVTEEPVPSLPDSYVIMELDKTKVKVGDIITAT 425
 QY 231 -----ETSLTVKETTETKETTETTTNNKQSTGCKET 262
 Db 426 IKIENMKNFAGYQLNIKVDPTMLEAIELETGSAIAKRTWPTVGTGIV-----LQSDNYGK 480
 QY 263 TS-----AKTQSIKTSKDLAPTSKVLAKPTPKAET 295
 Db 481 TAVANDVGAGIINFABAYNLTKYRTGVAEETGIIKIGIFRVLKAGSTAI-----RFD 535
 QY 296 TT-----KGPALTTP-----KPTPT-TPKEPASTTPKEP 324
 Db 536 TTAMPAIEGTWFDWYGENIKGVSVQPGIIVAEGEETPEPTVDTPTVTEEP 595
 QY 325 TPT----- 327
 Db 596 VPSELPDSYVIMELDKTKVKEGDIITIRVNNIKLAGYQIGIKYDPKLEAFNIETGD 655
 QY 328 ----- 327
 Db 656 PIDEGTWPAVGGTILKRDYLTGVAINNVSKILNFAAYVYVFDYEEGSEDTGIIG 715
 QY 328 -----TIKSAPTTPK-----EPATPTTKSAPTTPK 352
 Db 716 NIGFRVLKAEDTIRFEELSMPSGSDGTMLDWNLNRSIGVYVQIPAPKAAS-----D 770
 QY 353 EPATPTTKEPATTPKEPAP-----TTKEPAPTTSAPTTPKEPAPTTPKEPAPTTPKE 408
 Db 771 EPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 830
 QY 409 PAPT-----TPK-EPTPT-TPKEPAPT-TKEPAPTTPKEPAPTTPKPA--PTTPK 454
 Db 831 PPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD 890
 QY 455 EPAPT-TPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 513
 Db 891 EPTPSPTPEPTPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 947
 QY 514 TKEPAPTTPKEPAPTTPKPA-----PTTPKEPAPT-TPKEPAPTTPK 558
 Db 948 DTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD 1007

QY	559	APTAPKEPATTTPKEATPTTKLTPTPBKLAPTTPEKEPAPT---	TPBELAPTTPEEPT	615
Db	1008	EPTSPDEPTSD-----EPTSPDEPTSDDEPTSPD-EPTSPDEPT		1046
QY	616	PT--TPPEPAPTTPKAAAPNTKPEPAPTTPKEPA--PTTPKEPAPT--TPKETA	PTTPKGTGA	671
Db	1047	PSSTPEPIPTDTPSDEPTSDDEPTSDDEPTSDDEPTSEPIPTDTPSDE		1106
QY	672	PTTLKPEAPPTPKKAPKELAPTTTKEPT--STTSDDKAPATTPKGTATTPKEPAPTTPKE		730
Db	1107	PTPSDEPTPS--DEPTSD-EPTSPDEPTSETEEP IPTDTPSDEPTSDDEPTPS--DE		1161
QY	731	PAPTTPKGTATPTLKPEAPT--TPKKAPKELAPTTTKGPTSTTSDDKAPT--TPKETAP		786
Db	1162	PTPS-----DEPTSDDEPTSETEEPI-----PTDTPSDEPTSDDEPTSD-EP		1211
QY	787	TPKEPAPT--TPKKAPATTPETPTTSEVSTPTTKEPTTIHKSPDESTPELSABETPK		845
Db	1212	TPSDEPTSPETPEPIPTDTPSDEPTSD--EPTSPDEPT--PSDEPTP--SDEPTP-		1262
QY	846	ALENSKPEGPVPTTKPAATKPEMTTAKDKTTERDLRTTETTTAAPKMKETATTTTEK		905
Db	1263	--SETPEEP--IPTDTPSDEPTSDDEPTSDDEPTSDDEPTSDDEPTSDDEPTSPSE		1319
QY	906	TTSEKITATTQVSTTTQDTPPKITLTKTTLAPKVTTKKTIITTEIMNKPEETAKP		965
Db	1320	TPPEIPTDTPSDEPTSDDEPTSDDEPT-----PSDEPTP		1354
QY	966	KDRATNSKATTPXPO-KPTKAPKKPTSTKKPKMPRVKPKTTPTRKMTS-----		1015
Db	1355	SDPTSDDEPTSDDEPTSPSTPEPT-----PTTPTPTPTPTPTSGSGSG		1401
QY	1016	-----TWPELNPTSRIAEAMLTQTTTRPNQTPNSKLVEYNPKSEDAGGAEGE		1061
Db	1402	GSGGGGGGGGTVPSTPTPTS-----KPTSTPAPTIE-EPTPSDVPGAIGG		1449
QY	1062	TPHMLLRPH---VFMPB-----		1075
Db	1450	EHRAYLRGYPDGSFRPRBNITRAEAVIFAKLGADESYGAQSASPSYDLDTHWA		1509
QY	1076	-----VTPD-----MDYLPFRVNPQGI-----IINP		1095
Db	1510	KFATSOGLFKGYPDGTFKPDQNTTRAEBFATWJLHFLTKVKGQEIMSKLAID:SNP		1565
RESULT 14				
Q6SSE8	PRELIMINARY; PRT; 3889 AA.			
ID				
AC	Q6SSE8; 05-JUL-2004 (T-EMBLrel. 27, Created)			
DT	05-JUL-2004 (T-EMBLrel. 27, Last sequence update)			
DT	05-JUL-2004 (T-EMBLrel. 27, Last sequence update)			
DE	Minus agglutinin.			
GN	Name=SADI;			
OC	Chlamydomonas reinhardtii.			
OC	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;			
OC	Chlamydomonadaceae; Chlamydomonas.			
ON	NCBI_TaxID=3055;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CC-621;			
RA	Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,			
RL	Goodenough U.W.;			
RA	Submitted (Oct-2003) to the EMBL/GenBank/DBJ databases.			
RL	EMBL; AY450929; AAS07042.1; -			
DR	InterPro; IPR006315; Autotransporter.			
DR	InterPro; IPR008985; ConA-like lec.gl.			
DR	InterPro; IPR003882; Pistil_extensin.			
DR	PRINTS; PR01218; PSTLEXIENSIN.			
DR	TIGRFAMS; TIGR01414; autotrans_barl; 1.			
SQ	SEQUENCE 3889 AA; 389223 MW; 4DB5B4D5507214A CRC64;			

Db	2001	RISRYLRWSLLGIQGNIPLLDGAFFSSGSAAGSGSSSSNSSGGGLGDVDVAVALDRL	2061
Qy	975	--TTPKP-----QKP-----	982
Db	2061	QLSVPPPLPAAGDAASQAQPPANLSPPSASQLVADGSTALAGRRSRSLVQAAAPVAPSP	2120
Qy	983	--TKAPKP-----TSTKKPTMPR-VRKPKTTP-----	1008
Db	2121	PFTQAPTEPAFFTGAAPAPPAPPPQFPSTPPAPPMPALSLSGDRDLVLSWLOQVGA	2180
Qy	1009	-----TPKMTSTWPE-----LNP-----	1022
Db	2181	IGSGSNSSGGGSATSDSASVGAYSLGAAASPRGDVVVLDPGQLGLPGGSAPPMPQPG	2240
Qy	1023	-----TSRIEAWLQTTTRPN-----	1038
Db	2241	SGSGGQSGASVATDTTHTNVQDLLYLIALAALLMVALVAHLLVIGLWRLVAMVDVC	2300
Qy	1039	-----QTPNSKLV-----	1046
Db	2301	GAESGVEGHVLPFRPRAEWLGGLLVALTYALTUSGAASPRGNDTAAGRLIAVL	2360
Qy	1047	-----	1046
Db	2361	VLAVLVVPYGLLLWLTVCRWYLQEEVDHYMLGHPHQAFDGVIPGGAGAGDGGGHGA	2420
Qy	1047	-----EVPKSEDAGASEGT-----PHMLLRP 1069	
Db	2421	SALPAVAGFGTGGGGVAFACQAPBGGDGGGGEDGYGLGPHWALAP 2471	
RESULT 15			
AAS07042			
ID	AAS07042	PRELIMINARY; PRT; 3889 AA.	
AC	AAS07042;		
DT	02-MAR-2004 (TrEMBLrel. 27, Created)		
DT	02-MAR-2004 (TrEMBLrel. 27, Last sequence update)		
DT	02-MAR-2004 (TrEMBLrel. 27, Last annotation update)		
DE	Minus agglutinin.		
GN	SAD1.		
OS	Chlamydomonas reinhardtii.		
OC	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;		
OC	Chlamydomonadaceae; Chlamydomonas.		
OX	NCBI_TaxID=3055;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CC-621;		
RA	Ferris P.J.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,		
RA	Goodenough U.W.;		
RT	"Plus and Minus Sexual Agglutinins from Chlamydomonas reinhardtii.";		
RL	Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AY450929; AAS07042.1;		
SQ	SEQUENCE 3889 AA; 389223 MW; 4DE5B44D5507214A CRC64;		
Query Match 19.1%; Score 1124.7; DB 2; Length 3889;			
Best Local Similarity 13.6%; Pred. No. 2.1e-13;			
Matches 349; Conservative 153; Mismatches 446; Indels 1623; Gaps 63;			
Qy	1	MAWKTLPIVLLLLSVFVIOVSS-----QDLS-----	28
Db	22	LAFATL-----FFFLDLASTQTYGPNWATEDLNEQHGKLLAFILSGDTSEWSRP	71
Qy	29	SCAGRCGEG-----YSRATCNCDYNQHYNE-----CC-----	57
Db	72	EVATRLGFTGWRICNSCQTQFTSDQCAPDCESRTYCEPGAALSGENTTCALSLL	131
Qy	58	-----PDEKRVCT-----	65
Db	132	DQTYASQPPSSQPAWCSTYPGWGARPPRSVCDNFVFAARGTTPAGADDFDLAVSCSS	191
Qy	66	-----	65
Db	192	GTVPTYISGTRYRNDTVYRIMHNGAVTNPANVTRNOVKSIKLRSAWVHHPVSNITSP	251


```

QY 66 -----AEISC-----KGRCF-----75
Db 252 PEFSLVSELACPLESIFLEDVALRADYSVLMAQNTTLDNGKVFNLDPSTYFNVTGL 311
QY 76 -----ESPERG-----REC-----D 85
Db 312 YNVFVSMLEPSPMLHSSVSVKASSFNFGALRNAALRTCNFAPFTELHAMEWTQOAUQD 371
QY 86 CDAQ-----89
Db 372 WSAEYQAALSYVDGSQLRGRWLPPVIBPVNSARLLLPOTLARLTIIRTRDRQHGAVQWT 431
QY 90 -----CK 91
Db 432 SRPLITGLPGEWALLRNLEYLDLSDMETGAIVGPIPTWLMMSHLRVINMTGHNFCKR 491
QY 92 KYDK-----95
Db 492 DWHKILISQIRMYRAATHEPNLNVPHYYGFGNGNGMTRYNISVDYLSHGHWQYDEV 551
QY 96 -----CC-----PD-----Y-ESFC 104
Db 552 TTEAGFVEVIAPHGQCCKWKSQTIKDNVVEILYPGSRFGNNVQDEIYGGFYQDEWC 611
QY 105 AEVHNPTS-----PSSKKAP-----120
Db 612 -----EFTSPQPPPPAPSPSPPTTDPVPMPPSSPPAPVMPAPPQPPIPPASPLT 667
QY 121 -----120
Db 668 PAAAPRPLPTWPGKWEAGWPRPPIPRPRPRPPPLPPSPPLPVPTTSPSPRPPPK 727
QY 121 -----120
Db 728 SPPPPKSPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPSP 787
QY 121 -----PSSGA--SQTIKSTTKRSPKPNKKTKV-----146
Db 788 PPSPEPPSPAPSAAPSPSPSPAPSPDPSPKPPSPVPPSPPLPPSPSPVPSPPP 847
QY 149 ISEEEITEKVKONKQRTKKTKPPVVDVDEAGSLDNGDKVTTPTDSTTTOHNVSTSP 208
Db 848 PASPEPTSPAPSPPPPPSPSPAPSP-----SPPPSPSPSPAPSP 891
QY 209 KITTAKPINRPSLPSNDSKETSITVKNKTTVETKETTITNKQTSIDGKETTSAKET 268
Db 892 PPSPEPPSPAPPLPP-----907
QY 269 QIEKTSAKDLAPTSKVLAKTPKAEITTKGPALTTPKEPTTTPKEPASTTPKEPTPT 328
Db 908 ----PPSPPPSPAPSPPPSPSPAPSPPPSPSPAPSPPPSPAPSPPPSPAP 962
QY 329 IKSAPTTKEPAPTTKSAPTTKKEPAPTTKEPAPTTKEPAPTTKEPAPTTKSAPT 388
Db 963 PLPPPPSPSPAPSPPPPPSPAPLPPPPSPSPAPSPPPSPAPSPPPSPSPAP 1022
QY 389 TKKEPAPTTKKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTA 444
Db 1023 SPSPSP-EPSPAPSPPPSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP-A 1080
QY 445 PKKAPATTK--EPATTKEPAP-----TTTKEPSTTPKEPAPTTTSAPTTTKEPAPT 498
Db 1081 PPSAPSPSPAPSPAPSPPPSPSPAPSPPPSPSPAPSPPPSPAPSPPPSPAP 1140
QY 499 TTKSAPTTKEPSTTTKEPAPTTKEPAPTTKEPAPTTKEPAPTT-----TPKEPATTKEPAP 552
Db 1141 PPSPPPPSPSPAPSP-----PPSPSPSPSPPPPPSPSPPPSPSPAPSPAPPLPLPSP 1198
QY 553 TTTKAPATAPKEPAPTTKAPTTPKLTTPTEKLAPTTPKEPAPTTPELAPTPE 612
Db 1199 HTQSPSP-VPSPAPSPAPSPPPSPAPSPAPSP-APQAPSPSPFP-PPQPTAPTAP 1255

```

```

QY 613 EPTTTTPEEPATTPK-----AAANTPKAPATTTPKAPATTTPKAPATTTPK 668
Db 1256 PFPSPSPAPSPPTTSPSPAPQPPSTPHAPPSPSP-TPSPPLPPSPSPSP- 1313
QY 669 GTAPTTLKEPAPTTKAPKELAPTTTKEPTSTTSKAPATTTPKGTAPTTTPK 728
Db 1314 -----SPAPSPSPSP-----APPSPMPPS-----PAPLAPQSPSP-TPSPAPFPVP 1355
QY 729 KEPAATTTPKGTAPTTLKEPAPTTTPKAPKELAPTTTGTSTTSKAPATTTPK 788
Db 1356 PSPFPVPFPDPDPLPSPPTPSPFPVPFPSPPTPSPQPPS-----PAPSP- 1406
QY 789 PKEPAPTTK-----KPAATTTPETPPTTSEVSTTTTKEPTTIHKSPDESTPELSA 844
Db 1407 PLOPSPPPSPQPPSPAPGPPSP-----SPSTSPSPAPLAPAPPVPPMAPQPPS 1460
QY 845 KALENSKEPQVPTTKTAAKPE-----868
Db 1461 PPLSPFPFPPOPSPPTTIPASQPAPAAAALDCSAAATRTSFVASSSRGAFYIAVAPAS 1520
QY 869 -----868
Db 1521 SPSPYQVCGCCELSYAVLDPGASQOQVIPSQSSSSTAGGSPTVAVTSSVTPAGAGGLNG 1580
QY 869 -----MTTAKDKTTERDLR-----883
Db 1591 THGSTARRRALLVEATASSGPAAGVGARHLLLAATANSTTLEGLLATGRSRSAAGAGMG 1640
QY 884 -----TTPETT-----889
Db 1641 MSRLQVVDVQDGLDPVTAAPPTGTSNTTSGAGEAGSGTVRYSSMGAGSGGLDAWR 1700
QY 890 -----889
Db 1701 LTPGATGYLLRLKVADQEWWRWVSVIDPRAAGQLLLARRTGGSSSSNSTSGSALA 1760
QY 890 -----889
Db 1761 AADEVQVENVNAAAGATAVSAQAAVRMLMAVIAVSEFVQPFSTLSALRSGGARLL 1820
QY 890 -----TAAPKM-----TKETATTEKTTESKITATTQVTTSTTT-- 923
Db 1821 STQCFASATAAAEVAAGTVDASTPPGSDASSAATVAPAAIAPVSGTSGT--SNS 1880
QY 924 -----QDTTPKLTITLKT-----LAPKVTITTKTITTIMN 957
Db 1881 SAYQSCVAVLFAEQDAPTELLUPPGTLTDMHGNINAEPLILSVNLTASADSUSTVERAG 1940
QY 958 KPEETA-----KPKDR 968
Db 1941 APVAAAVAGGVFAASAFTSASAPLSAFSSRSLLQSGYHIQMLAMSSSLASPGISAFR 2000
QY 969 -----ATNSKA-----974
Db 2001 RISYLRWSLLGIQCNIPLLDGFSGSAAAGSGSSSSSSSSGGLGDVDAVALDRL 2060
QY 975 --TTPK-----QKP-----982
Db 2061 QLSVPPPLPAAGDAASQAQPPANLSPPSASOLVADOSTALAGRSRSLVQAARVAPSP 2120
QY 983 --TKAPKKP-----TSTKKPKTMR-VRKPKTTP-----1008
Db 2121 PFTQAPTAPFGTGTGAAPAPPAPPPQSTPPPPAPPMPEALSLSGDRDLVSLWMLQOVGA 2180
QY 1009 -----TPRKMTSTPE-----LNP-----1022
Db 2181 IGSNSSSGGGSATSDASVAYSLOGGAAAPRGDVVLPDGGQLLLGPGGSAPPMPQPG 2240
QY 1023 -----TSRIAEAMLOTTTRN-----1038
Db 2241 SSGGQSGSGASVATDTIHTNVQDILYTLIAALLMVALVAAHLLVIGLRLAVMYDVC 2300
QY 1039 -----QTENSKLV-----1046

```

```
Db      2301 GAAESGVEGLHPEVLRPPRAEMVLGGLLLVALTFYSALTLGAAEPRWGDNTAAGRLIAVL 2360
Qy      1047 ----- 1046
Db      2361 VLAVLVVPYGLLLWLTVCRWYLQEEVDHYMLGPHWQAFDQVIPGGAGAGDGGGHGA 2420
Qy      1047 -----PHMLLRP 1069
Db      2421 SALPAVAGFGTGGGGVAFACQAEPSGGGGGGEDGYGLGPHWALAP 2471
```

Search completed: October 13, 2004, 12:17:37
Job time : 156.725 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 22.3651 Seconds
(without alignments)
4723.689 Million cell updates/sec

Title: SEQ1-C
Perfect score: 5887
Sequence: 1 MAWKLPYVLLLSVFVIQ.....DMDYLPVNPQGIINPMLS 1098
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1325	22.5	3020	2 A43932	mucin 2 precursor,
2	1132	19.2	1664	2 T18262	S-layer protein -
3	1020.9	17.3	1889	2 T31108	cyst germination s
4	1010.7	17.2	7862	2 I38346	elastic titin - hu
5	1008.3	17.1	1274	2 T16251	hypothetical prote
6	996.8	16.9	2187	2 T30826	nascent polypeptid
7	992.1	16.9	3570	2 T45025	mucin MUC5B, trach
8	923.9	15.7	1367	1 S49478	glucan 1,4-alpha-g
9	910.5	15.5	3507	2 T34513	hypothetical prote
10	903.1	15.3	1188	2 S49915	extensin-like prot
11	872.9	14.8	6842	2 T29757	protein UNC-99 - C
12	871.5	14.8	1229	2 T25697	hypothetical prote
13	861.9	14.6	1151	2 T18535	high molecular mas
14	858.5	14.6	2897	2 B48666	cell proliferation
15	855.1	14.5	3256	2 A48666	cell proliferation
16	840.3	14.3	4135	2 T42629	tenascin-X - bovin
17	835.2	14.2	5762	2 A41819	proline-rich pepti
18	827.5	14.1	3341	1 A35175	mucin 1 precursor,
19	783.8	13.3	2232	2 T34434	hypothetical prote
20	778	13.2	2142	2 B35098	MHC class III hist
21	777.2	13.2	5262	2 T31113	ALR protein - huma
22	767.9	13.0	1832	2 T31113	mucin-like glycopr
23	766.1	13.0	3942	2 T42730	Bassoon protein -
24	765.9	13.0	590	2 I51618	nucleolar phosphop
25	765.4	13.0	761	2 T84672	hypothetical prote
26	760.9	12.9	4006	2 T03070	probable tenascin
27	751.9	12.8	1870	2 S43751	MHC class III hist
28	751.2	12.8	2774	2 A43359	microtubule-associ
29	750.9	12.8	1872	2 S35152	MHC class III hist

30 749.5 12.7 2225 2 T26063
31 747.7 12.7 4548 1 S00657
32 745.7 12.7 924 2 S27923
33 742.3 12.6 3938 2 T42761
34 732.2 12.4 971 2 T19431
35 723.6 12.3 801 2 T29018
36 723 12.3 1777 2 T34369
37 722.9 12.3 4957 2 T03455
38 722.3 12.3 3381 2 T42389
39 718.2 12.2 1611 2 T38236
40 715 12.1 1630 2 A53577
41 703.7 12.0 3566 1 A40701
42 702.8 11.9 1952 2 T48814
43 700.7 11.9 5170 2 T15348
44 696.5 11.8 3968 2 A44265
45 696.2 11.8 3869 2 A48205

ALIGNMENTS

RESULT 1

A43932
mucin 2 precursor, intestinal - human (fragments)
N/Alternate names: mucin SMUC-41
C/Species: Homo sapiens (man)
C/Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text change 09-Jul-2004
C/Accession: A49963; A45106; A43932; B35332; A61257; PQ0328; PQ0329
R/Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A/Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A/Reference number: A49963; MUID:94132002; PMID:8300571
A/Accession: A49963
A/Molecule type: mRNA
A/Residues: 1-639 <GU1>
A/Cross-references: UNIPROT:Q02817; UNIPROT:Q14884; GS:L21998
R/Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A/Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr:
A/Reference number: A45106; MUID:93015075; PMID:1400449
A/Accession: A45106
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 626-1895 <GU2>
A/Cross-references: GB:M94131; NID:G186395; PIDN:AAA59163.1; PID:G186396
A/Note: sequence extracted from NCBI backbone (NCBIP:116706)
A/Accession: B45106
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 2037-3020 <GU3>
A/Cross-references: GB:M94132; NID:G186397; PIDN:AAA59164.1; PID:G186398
A/Experimental source: colon
A/Note: sequence extracted from NCBI backbone (NCBIP:116698)
R/Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.I
J. Clin. Invest. 88, 1005-1013, 1991
A/Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymor:
A/Reference number: A43932; MUID:91358717; PMID:1885763
A/Accession: A43932
A/Molecule type: DNA
A/Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A/Cross-references: GB:M74027; NID:G188863; PIDN:AAA59875.1; PID:G188864
A/Note: sequence inconsistent with the nucleotide translation
A/Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIP:55750)
R/Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A/Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden:
A/Reference number: A33532; MUID:89197956; PMID:2703501
A/Accession: B33532
A/Molecule type: mRNA
A/Residues: 1916-2193 <GU4>
A/Cross-references: GB:M22405; NID:G188873; PIDN:AAA36334.1; PID:G188874
A/Experimental source: intestine
R/Jary, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

J. Clin. Invest. 87, 77-82, 1991
 A:Title: Human bronchus and intestine express the same mucin gene.
 A:Reference number: A61257; MUID:91086481; PMID:1985113
 A:Accession: A61257
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1,1925-1948, 'TTS', 1952-1954 <JAN>
 A:Experimental source: bronchus
 R:Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Jones, C.; Forstner, Biochem. Biophys. Res. Commun. 183, 821-828, 1992
 A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-b
 A:Reference number: PQ0328; MUID:92198477; PMID:1550588
 A:Accession: PQ0328
 A:Molecule type: protein
 A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
 C:Genetics:
 A:Gene: GDB:MUC2
 A:Cross-references: GDB:120203; OMIM:158370
 A:Map position: 11p15.5
 C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von Willebrand factor type C repeat homology
 C:Keywords: glycoprotein; intestine; tandem repeat
 F:2766-2834/Domain: von Willebrand factor type C repeat homology <VMC>
 Query Match 22.5%; Score 1325; DB 2; Length 3020;
 Best Local Similarity 20.8%; Pred. No. 1.5e-18;
 Matches 417; Conservative 102; Mismatches 500; Indels 990; Gaps 64;

QY	6	LPYILLVSVFVIOQVSODLSSCAGRCG-----EG-----YSR 40
DB	502	LQVLAPVMQLEVTLDQASQ--GQVGLCGNPNGLGDDFKTAGLVEATGAGFANWKA 559
QY	41	DATCN-----CDY---48
DB	560	QSTCHDKLDWDDPCSLANIESANYAEHWSLLKKTETPFGRCHSAVDPAEYVKCKYDTC 619
QY	49	NOHYWEC---CPDFKRVCTA-----ELSK 71
DB	620	NCQNNEDCLCALSSYACTAGKVMWGVRHVCNKGDCVGNPSQVFLYNLTTCQOTCR 679
QY	72	-----GRCFESF-----ERGR-----ECDC-----DAQCKYDK 95
DB	680	SUSEADSHCLEGFAPVDGCGCPDHTFLDEKGRCVPLAKSCYHRLGLYLEAGDVVVROER 739
QY	96	C-----96
DB	740	CVCRDGLHCRQIRLIGQSTAPKIHMDCSNLATATSKPRALSCQTLAGYHTECVSG 799
QY	97	---CPD-----YESFCAEVHNPSPSSKKAP-----120
DB	800	CVCPLMDDDRGCGVKECEPCVHNNDLYSSGAKIKVDCNTCTCKRGRWVCTQAVCHGT 859
QY	121	-----PPSG-----ASQT 128
DB	860	CSIGSGHVIITFDGKYVDFDGHCSYVAVQDYCGQNSLSGSFIITENVPCGTTGVTCSKA 919
QY	129	IKSTTKRSPKPNKKTKKVIIESE-----ITEKVKNNKRNKKKTPKPPVVDGAS 182
DB	920	IKIFMGRTELEKLEDKHRVIOQDEGHVAYTTREVGQY-----LVVESST 964
QY	183	GL-----DNGDFKV-----191
DB	965	GIIVWDRKTTVFIKLAPSYKVTGCLGNGFDRHSNDDFTTRDMVVSSELDFGNSWKEA 1024
QY	192	-TPDTSITQ-----HNKVSIP-----208
DB	1025	PTCPDVSTNPEPCSLNPHRRSWAEKQCSILKSSVFSICHKSVDPKPFYEAACHDSCSDT 1084
QY	209	-----KITT 212

DB	1085	GGDCEFCFSAVAYAEQCTKEGACVFWRTDLCPICDYVNPPEHECEWHYBPCGNRSFET 1144
QY	213	AKPIN-----PRPSLP- 223
DB	1145	CRTINGIHSNISVSVLEGYPCPCPKDRPIYEBDLKCVTADKCGCVEDTHYPGASVPT 1204
QY	224	-----PNSDTSKE-----TSLTVNKET 240
DB	1205	EETKSCVCTNSQVVRPEEGKILNQDQAGFCWE:CGPNGTVKEHFNCSITRPSST 1264
QY	241	-----TVETKRETTTINKQTSTDGKEKTSK-----266
DB	1265	LTFVTITLPTPTPTSTTTTTTTTTTSSVLSSTFKLCLMSDWINEDHPSSGSDGDRE 1324
QY	267	-----ETQSIKTSKAD-----278
DB	1325	PFDGVCAPEDIECRSVKDPHLSLEQHGQKQVQCDVSGVFIKCNEDQFGNGPFGLCYDKI 1384
QY	279	-----LAPTSKVLAKPTPKAETTKGPAITTKPEPTTTPKBPASTTKPEPTPT 327
DB	1385	RVNCWPMKCIITTPSPPTTTTULPPTTTPSPPTTTTTPPTTTPPTTTPPTTTPPTT 1444
QY	328	TIKSAP-TTPKBPAPTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 386
DB	1445	TTTTLPTTTPSPPTTTPPTTTPSP-PTTTPSPPTTTPSPPT-TTTTTPPTTTPSP 1502
QY	387	P-TTP-----KEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 437
DB	1503	PMTPITTPSPPTTTPPTTTPSPPTTTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTT 1562
QY	438	KEPAPTKKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 497
DB	1563	SPPTTTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1621
QY	498	TTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 556
DB	1622	TTT---PITP---PTSTTTLPPTTTPSPPTTTTTPPTTTPSPPTTTTTPPTTTPSPPTTTTTPPTT 1674
QY	557	KPAPTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 616
DB	1675	PPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1725
QY	617	TTPEEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 664
DB	1726	TTMTTTPSPPTTTPSPPTTTTTPPTTTPSPPTTTTTPPTTTPSPPTTTTTPPTTTPSPPTTTTTPPTT 1778
QY	665	TTT-----KG-----669
DB	1779	TTPCVPLCNWGTGMDSGKPNFHKPGDTELGVCVPGWAAANISCRATWYDPVPIGQLGQ 1838
QY	670	-----TAPTTLKEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 688
DB	1839	TVVCDVSVGLICKNEDQKPGGVIWMAFCLNYBINVQCCECVTQPTM---TTTTTENPTP 1895
QY	689	KELAPT--TKPSTSTSDK-----APT--TPKGTAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 741
DB	1896	TPITTTTTPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1954
QY	742	TTLKEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 800
DB	1955	TVPTPTPTGCTPTTTLTITTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 2014
QY	801	-----APTPPTPTPTTSEVSTPTTK-----EPTTIHKSDESTPELSREPTPKALE 848
DB	2015	TMIPISITTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 2074
QY	849	NSPKBPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 908
DB	2075	STTLLSTLPAEMTSTAPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 2132
QY	909	SKITATTQTSTTQDTPFKITT---LKTTLAPKVTTKTITTTTTEIMNKPBE---961

Db 2133 SAPTPSTVTTTSAWPTPTPLSTPSIIIRTTGLRPPSSVLICCVLNDYVYAPGEVYN 2192
QY 962 -----TAKPKDRATNSKATTPKPKQPTKAPKPTSTKKPK 996
Db 2193 GTYGDTCYFVNCSLCTLEFYNWSCPSPTSPPTPSK-STPTPSKSPSTPSPTGTGTPP 2251
QY 997 TMRVRKPKPTTPTPRKMTMTMELNPTSIAEAMLTQTTTRPNQTNPKSLVEVNP ----- 1050
Db 2252 ECPDFDPPR-----QENETWLCDFM-ATCKYNNVTVEIVKVECEPPMPTC 2297
QY 1051 -----KSEDAGGA-----EGETHMLLRPHVFMPE 1075
Db 2298 SNGLOPVRVEDDGCCHWNECDCTGWDGPHVTFDGLYSGYQNCYIVLVE-----E 2351
QY 1076 VTPTDMD-----YLPVRVNOGHIIN 1094
Db 2352 ISPSVDNFGVYIDNVHCDPNDKVSCTPLIVRHETQEVLIKTVHMPMQVQVQVNRQAVA 2411
QY 1082 -----YLPVRVNOGHIIN 1094
Db 2412 LPYKKGLEVYQSGINYYVDIPELGLVLS 2440
RESULT 2
Ti8262
S-layer protein - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18262
R:Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
e.
A:Reference number: Z18847; MUID:93209931; PMID:8458832
A:Accession: T18262
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1664 <FUU>
A:Cross-references: UNIPROT:Q06852; EMBL:X67506; NID:G296879; PID:G296881; PIDN:CAA47841
Query Match 19.2%; Score 1132; DB 2; Length 1664;
Best Local Similarity 21.4%; Pred. No. 3.8e-15;
Matches 371; Conservative 124; Mismatches 421; Indels 820; Gaps 69;
QY 4 KTLPIVLLLLSVFVIQVSSODLSGACGCGEYSDATCNDVNCQHYMECCP----- 58
Db 6 KVLISILLTLL-----LISTSVMS-----PFAETPSIEMV 37
QY 59 -----DFKRVCTABLCKGRCFCFESFERGREGCCDCAQCKYDKCOPD 99
Db 38 LDKTEVHVGDVITATIKVNNIRKLAGYQLNIK----- 69
QY 100 YESFCAEYHNPTSP-----PSSKKAPPSPSGASQTIKS----- 131
Db 70 ---FDFEVLQVDPATGTEFTDKSPVNRVLLTNSKYGTPV-AGNDIKSGIINFATGYN 125
QY 132 ---TKRSP-----KPNKKKTK----- 146
Db 126 NLTAAYKSGIDHTGIGIGIFKVLKQNTSIRFEDTILSMGALSGTSLFDWDAETITGY 185
QY 147 -----KVIESEETE-----KVK-----DNKN-----RTKKKPT- 171
Db 186 EVIQPLIVVEAPLKDASVALELDKTKVKYGDIIITATIKIENMKNFAGYQLNIKYDPTM 245
QY 172 -----PKPVDVEAGSLDN----- 186
Db 246 LEAIELETGSAKRTWPVTGGTVLQSDNYGKTTAVANDVGAGIINFAYEASNLTKVRET 305
QY 187 -----GPKVTTPTD-----STQHNKV 204
Db 306 GVAEETGIIGKIGFRVLKAGSTAIRFEDTTAMPGAIEGTYMFDWYGENIKGYSVVQGEI 365
QY 205 -----STSPKITTAKPINRPSLP-----PNS-----DTSK----- 230

Db 366 VAEGEPGEPTPEEPTETPTVTPTEEPVPSLPSVIMELDKTKVKVGDIIITAT 425
QY 231 -----ETSLTVNKEETTETKETTTTTNQTSTGDKCKT 262
Db 426 IKIENMKNFAGYQLNIKYDPTMLEAIELETGSAIAKRTWPVTGGTV-----LOSDNYGKT 480
QY 263 TS-----AKETQSIKTSIAKOLAPTSKVLAKPTPKAET 295
Db 481 TAVANDVGAGIINFAYEASNLTKYRETGVAEETGIIKIGFRVLKAGSTAL-----RFED 535
QY 296 TT-----KGPALTP-----KEPTPT-TPKEPASTTPKEP 324
Db 536 TTAMPGAIEGTYMFDWYGENIKGYSVVQGEI VAEGEPTPEEPTETPTVTPTEEP 595
QY 325 TPT----- 327
Db 596 VPSLPSVIMELDKTKVKEGDVIIATIRVNNIKNLAGYQIGIKYDKPVLEAFNIETGD 655
QY 328 ----- 327
Db 656 PIDEGTWPVGGTILKNRDYLP TGVAINNVS KGIILNPAAYVYVFDYRERKGSSEDTGIIG 715
QY 328 ---TIKSAPTTPK-----EPATTTKSAPTTPK 352
Db 716 NIGFRVLKASDTHIRPEELSMPCSGIDGTYNLWYLNRIISGYVVIQAPAKAAS-----D 770
QY 353 EPAPTTTKEPAPTTKPEPAP-----TTTKEPAPTTTKSAPTTPKPEPAPTTTKKAPATPKPE 408
Db 771 EPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPEEPIPTDTPSDEPTPSDE 830
QY 409 PAPT-----TPK-EPPT-TPKEPAPT-TKEPAPTTPKEPAPTAPKPA--PTTPK 454
Db 831 PTPSDEPTPSDEPTPSDEPTPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSD 890
QY 455 EPAPT-TPKEPAPTTTKEPAPTTKPEPAPTTKSAPTTPKPEPAPTTTKSAPTTPKESPT 513
Db 891 EPTSEPTPEEPIPTDTPSDEPTPSDEPTPS-----DEPTPSDEPTPSDEPTSEPEEPIPT 947
QY 514 TTKEPAPTTKPEPAPTTKPKA-----PTTKPEPAPT-TPKEPAPTTTKKP 558
Db 948 DTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPEEPIPTDTPSD 1007
QY 559 APTAPKEPAPTTPKETAPTTPKLTTPPEKLAPTTPEKAPT--TPSELAPTTPEEPT 615
Db 1008 EPTPSDEPTSD-----EPTPSDEPTPSDEPTPSD-EPTPSDEPT 1046
QY 616 PT-TPKEPAPTTKAAAPNTKPEPAPTTPKPA--PTTPKEPAPT-TPKETAPTTPKGT 671
Db 1047 PSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPEEPIPTDTPSDE 1106
QY 672 PTLKKEPAPTTKPKAPKELAPTTTKEPT-STTSKAPATTPKGTAPTPKEPAPTTPKE 730
Db 1107 PTPSDEPTPS--DEPTPSD-EPTPSDEPTPEEPIPTDTPSDEPTPSDEPTPS--DE 1161
QY 731 PAPTTPKGTAPTTLKPEPAPT-TPKKAPKELAPTTKGTSTTSKAPAPT--TPKETAP 786
Db 1162 PTPS-----DEPTPSDEPTPEEPI-----PTDTPSDEPTPSDEPTPSDEPTPSD-EP 1211
QY 787 TPKKEPAPT-TPKKAPATTPETPPPTSEVSTSTTTKEPTTHKSDSTPELSABPTPK 845
Db 1212 TPSDEPTPEEPIPTDTPSDEPTPSD--EPTPSDEPT--EPTPSDEPT--SDSPTP- 1262
QY 846 ALENSKPEGVPTTKTAPKPEMTTAKOKTTERLRTTPTTTTAAKPKMTATATTEK 905
Db 1263 --SETEEP-IPDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1319
QY 906 TTESKITATTTQVSTTTQDTPFKITLTKITTLAPKVTITTKITTTTEIMNKPEETAKP 965
Db 1320 TPEEPIPTDTPSDEPTPSDEPTPSDEPT-----PSDEPTP 1354
QY 966 KDRATNSKATTPKQ-KPTKAPKKPTSTKKPKTMPRVRKPTTPTPKMTS----- 1015

Db 1355 SDEPTPSDETPSDETPSETPBEPT-----PTTPTPTPTPTPTSGSGSG 1401
 QY 1016 -----TPELNTSRIAEAMLQTTTRPQNTSKLVBNPKSEADAGGAEGE 1061
 Db 1402 GSGGGGGGGGTVPTSPPTPTS-----KPTSTPAPTLE-EPPTSDVPGAIGG 1449
 QY 1062 TPHMLLRPH---VMNPE----- 1075
 Db 1450 EHRAYLRGPDGSGFRPERNITRAEAAVIFAKILGADESYGQAASPYSDLDATHAAWAI 1509
 QY 1076 -----VTPD-----MDYLPVPPNQGI-----IINP 1095
 Db 1510 KPATSQGLFKGYDPDTPKPDQONITRAEPATVVLHLFKVKQBIWMSKLATIDISNP 1565

 RESULT 3
 T31108
 Cyst germination specific acidic repeat protein precursor - Phytophthora infestans
 C;Species: Phytophthora infestans (potato late blight agent)
 C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
 C;Accession: T31108
 R;Goernhardt, B.
 submitted to the EMBL Data Library, April 1998
 A;Reference number: Z20986
 A;Accession: T31108
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-1489 <GOE>
 A;Cross-references: UNIPROT:O96449; EMBL:AF061185; NID:g3851513; PID:g3851514; PIDN:AACT
 C;Genetics:
 A;Gene: car90

 Query Match 17.3%; Score 1020.9; DB 2; Length 1489;
 Best Local Similarity 26.1%; Pred. No. 4.9e-13;
 Matches 406; Conservative 62; Mismatches 425; Indels 661; Gaps 74;

 QY 41 DATCNDCNQCQHYMEC-----CPDFKRVCTAELSCKGRCFBSFERGRCDCDAQ 89
 Db 32 DATYLSLSSCG--IFCSGVGAEPVGTACPKAGDVATS-----DCQPY 71
 QY 90 CKYPD--KCCPYESFCVAEN-----PTSPSSKKA----- 119
 Db 72 LLSYNGAVCVAPDAECALIHDDMMGCEFPKGTGYTSAVEAETIAAYNGESSGWTGHDEV 131
 QY 120 -----PPPSG----- 124
 Db 132 VQVGDEBEI PARVNYDVTDTPIGVNCEVATETATQHAEGGKYVDTPSTGTQTDYG 191
 QY 125 -----ASQTIKSTKRSKPKPNKKTKKVISSSE-----ITERKVDNKNRKKKPTTP 172
 Db 192 NTHYGSSTTEGVTGGYGPTD-----AKVIDGETVLDYPTGITIELID----- 235
 QY 173 KPVVDKAGSLDNGDFKVTPTDSTT-----QHNKYSTSKI 210
 Db 236 -----GTPGGYGTITDGGTITGCGYTVDNTHETGAGGYDAGTBEESTPTFV 285
 QY 211 -----TTAKPIN-----PRPSLP----- 223
 Db 286 GYSTETEQQVTGGVPSDETEAFTGTYVVPRETTAAPSEDITVAPREVTPVAPTEK 345
 QY 224 -----PNSDT----- 228
 Db 346 PYDVBETTVYBEETVAPTKSTNAPTERMYAHIEKPCDTEVVMYAPTEETVAPTEET 405
 QY 229 -----SKETSLTVNKKETIVE-TKETTT-TTNKQSTSDGKEKTSKAKTSQIEKTSKADLAPT 282
 Db 406 TVAPTEETVYATEETPYEPTTEETTVTPTEETVYAPTEET-----YAPT 450
 QY 283 SKLAKPTPKAETTTKGPALTKPKPTPT-----PKE 315
 Db 451 EKTTVAPT---EETTVAPTEETPYEPTTEETTVYAPTKETTVAPTEETVYASTEETVAPTE 507
 QY 316 -----PASTPKPEPTPTTIKSAPTTPKEPAPT-TTKSAPTTPKEPAPT--TTKEPAPTTP 367

RESULT 4

I38346

elastic titin - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004

Db 508 ETTYAPAEETPYEPTTEET--TVAPTEETVYAPTEETVYAPTEETVYAPAEETP 566
 QY 368 KBEPAPTTTKEPAPTTTKSGAPT-----TPKEPAPTTTTPKAPPTTPKAPPTTPKEP 417
 Db 567 YETPEETTVYAPTEET--YAPTETVMYAPIETTVYAPTEETVYAPAEETPYEPTTEETVAP 625
 QY 418 TPTTTPKEP-----AP---TTKEPAPTTPKEP-----APTAKKAPAPTTPK 454
 Db 626 TBEETTVYAPTEETVYASTEETTVYAPTEETVYAPAEETPYEPTTEETVYAPTEETVAPTEET 685
 QY 455 EBAPTTPKEPAPT--TTKEPSPPTTPKEP-----APT-TTKSAPT-----TTKEP 495
 Db 686 TVAPTEETVYASTEETTVYAPAEETPYEPTTEETVYAPTEETVYAPTEETVMYAPIETTVYGP 745
 QY 496 -----APT-TTKSAPT--TPKEPSPPTTPKEP-----APTTPKEPAPTTPKAPAPTTPK 540
 Db 746 TBEETTVYAPTEETVYAPTEETVYAPTEETVYETPTTGTETVYAPTEETVYAPTEETVYAPTEET 805
 QY 541 EPAPT--TPKEPAPTTTKKAPAPTAPKEP-----APTTKETAPTTPKLTTP 584
 Db 806 TVAPTEETPYEPTTEETTVAPTEETPYEPTTEETTVYPTTEETVYAPTEETVYAPTEETVAP 865
 QY 585 TTPEKLAPT--TPKEP-----APTTPEBELAPT-----T 610
 Db 866 TBEETTVAPTEETPYEPTTEETTVAPTKETTVYAPTEETTVASTTEETVYAPTEETVYAPAEET 925
 QY 611 PEPPTTPTEEP-----APTTPKAAANTPKAPAPT--TPKEPAPTTPKEP-----AP 656
 Db 926 PYEPTTEETVYAPTEETVYAPTEETVYAPTEETVYAPTEETVYAPAEETPYEPTTEETVAP 985
 QY 657 TTPKETAPTTPKGTAP-----TTLKEPAPTTPKKP-----APKE-----LAPT--- 694
 Db 986 TBEETTVAPTEETVMYAPIETTVYAPTEETVYAPAEETPYEPTTEETVYAPTEETVYAPTEET 1045
 QY 695 -----TTKETSTITSKPA-----PTTPKGTAPTTPKEPAPTTPKAPAPTTPKGTAP 741
 Db 1046 TVASTEETVYAPTEETVYAPAEETPYEPTTEETVYAPTEETVYAPTEETVYAPTEETVYAP 1105
 QY 742 T--TLKEPAPTTPKKP-----APKE-----LAPT-----TTKGPTSTTSKAPAPT 780
 Db 1106 TBEETTVAPAEETPYEPTTEETTVYAPTEETVYAPTEETVMYAPIETTVYAPTEET--YAPTE 1163
 QY 781 PKETAPT--TPKEPAPTTPKKP-----APTTPEPTPPPTTSEVSTPT--TTKEPTTIHKS 830
 Db 1164 ATTVAPTEETVYAPTEETVYETPTTGTETVYAPTEETVYAPTEETVYAPTEETVYAPT--EET 1221
 QY 831 PDBSTBELSAEPTPKALENSPKPEGVPTTKTPAATKPEMTTAKDKTTERDLRTTPTETT 890
 Db 1222 PYEPTTEETVYAPT-----EETVYPTTEETVYAP-----TEETVYAPTEET 1262
 QY 891 AAPKMTKET-----ATT--TEKTESKITATTQVSTTT--TQDPT--PFKITT-- 933
 Db 1263 YAP--TEETVMYAPIDETTVYPTTEETVYAPTEATTVAPTEETVYAPTEETVYETPTGETTVYA 1320
 QY 934 -LKTITLAKVITT-----KKTITITIMNKPEETAKPKDRATNSKATTP 977
 Db 1321 PTEETTVYAPTEETVYAPMEETVYEPABESTSTVSTKPCNTEETTEDEPTDEPTDPSDEP 1380
 QY 978 KPQ---KPTKAPKKPTSTKKKXTMPRVKRP-----MTSTMPELNPTSRIAEAMLQTTRENQTPN 1011
 Db 1381 TDEPTDEPTDLP-----TDEPST-----PCDNOGINGIGVENKVRNAGIYNTTPGPR 1429
 QY 1012 K-----MTSTMPELNPTSRIAEAMLQTTRENQTPN 1042
 Db 1430 NSQSWHSCCRSCVNDPICHAFPHOTS-----SDSVCELTTSTSDREEDQON 1476

C;Accession: I38346
R;Label: S.; Kolmerer, B.
Science 270, 293-296, 1995
A;Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A;Reference number: A57430; MUID:56026330; PMID:7569978
A;Accession: I38346
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-7962 <RES>
A;Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:gl017426; PIDN:CAA63189.1; PID:gl017426
C;Genetics:
A;Gene: GDB:TTN
A;Cross-references: GDB:I27867; OMIM:188940
A;Map position: 2q31-2q31

Query Match	17.23;	Score	1010.7;	DB 2;	Length	7962;			
Best Local Similarity	14.98;	Pred. No.	7.7e-12;						
Matches	390;	Conservative	121;	Mismatches	447;	Indels	1653;	Caps	94;
Qy	4	KTLPYILLLLSVFVIQOVSSQDLSSCAGRCGEGYGRDATCNDY	----	DNDVAFEDIKINYPKLSWKGT	5359				
Db	5317	QTEPVTL	----	IKDINQTVLK	----	48			
Qy	49	-----VCT	----	NCQ	----	62			
Db	5360	EKLEPSDKFEISIDGRHTLRVKNCOLKQOQNYRLVCGPHIASAKLTVIEPAWERHLODV	----	HYMECC	----	5419			
Qy	63	-----VCT	----	ASL	68				
Db	5420	TLREGQCTWTVQFSPNVKSEFRNGRILKPGQRHKTVEVHKVHKLTADVRAEDQGV	5479						
Qy	69	SKGRGFESFEREGECDDAOCKKYDK	----	CCPDYESPCAE	----	106			
Db	5480	TCK	----	YEDLETSAELEIAEPIQFTKRIQNIIVVSEHQSATFCEVSFDDAIVTWYKGT	5536				
Qy	107	-----VHN	----		109				
Db	5537	ELTESQKYNFRDGRCHYMTIHNVTDPDEGVSVIARLEPRGEARSTABLYLTKEIKLE	5596						
Qy	110	-----PTSP	----	PSSKKAPP	121				
Db	5597	LKPPDIPDSRVPIETMPIRAVPPEIIPVVPVAPVPLLLFTPEBKPPPKRIEVTKAVKK	5656						
Qy	122	-----PSGASQTI	----		129				
Db	5657	DAKKVAKPKEMTBEIRIVKKPPPTTLIPAKAPEIIDVSSKAEVYKIMTIRKKEVQKE	5716						
Qy	130	-----	----		129				
Db	5717	KEAVYKKQAVHKEKRVFTIESFEPEYDELEVEPYTEPFQPYVEEDPEDEYEIKVEAKKE	5776						
Qy	130	-----KSTT	133						
Db	5777	VHEEWEDPEEGQYVEREBGYDEGEEMWEEAYQEREVIQVQKEVYVESHERKVPKVPFE	5836						
Qy	134	KRSPKPPN--KK	----	KTKKVISSB	----	166			
Db	5837	KKAPPPPKVKKPVEIKI EKTSRMMEEXQVQTVKPEVSKKIIVQPKPSRTPVQEEVIEVK	5896						
Qy	167	-----	----		166				
Db	5897	VPVHTKQWVISEBKMFASHTEBEEVSVTVPEVQKEIVTEEKIHVAVSKRVBPFPKVPPL	5956						
Qy	167	-----KKKPTXP	----		175				
Db	5957	PEKPAPPEVAPVPIPKKVEPPAPKVPBPVPEEKKVPVPPKKEPAPPAAPPKVPVPPKPF	6016						
Qy	176	VVDAGSLDNGDFKVTTPDTSITQ	----	HNKYSTSPKITT	----	213			
Db	6017	VPZE	----	KLPVPVAKKKEAPPKVPVQVGVTEKIIITVQRESPPPAPVE	6066				
Qy	214	-----KPINPR	----	PSLP	----	231			

QY 754 KPAPKELAPTTTKGPTSTSDKAPATTPKETAATTPK----- 790
: : : : :
Db 1022 DSPSKKAAPVAA-----KPDPKIP-EVSPPTPVKNFVKWKPWEDDEPSPVSGAP 1071
QY 791 EPATTP--KKDAPTPPEPPT--TSVSTPTTKETTHKSPDESTPESLAEPTPK 845
: : : : :
Db 1072 BEKKTIPVLAKAPTKAPKPSAAAADPVSGP-TSKDPLKSKAPVEK-PKPTTDPKDD 1129
QY 846 ALSENSPKPGVPTTKPAATKPEMTTAKDKTTERDLRTTPTTAAAKMTKETAITTEK 905
: : : : :
Db 1130 KLPSP-----AKPE-----KAPE--PAAPKKWKPWDDDDPD 1160
QY 906 TTESKITATTQVSTTTQDTTPFKITTLTKTTLAPKVTTKTITTEIMNKPSETAKP 965
: : : : :
Db 1161 EPEADTVAPSKKPTEDPADPLG-----GPKTKDPK-----LNKKAPAEKP 1203
QY 966 KDRATNSKATTPKPKQKPTKAPKKPTSKPKTMPVRVKPK-----TTPTPRKM 1013
: : : : :
Db 1204 TEK-----PKPEVSKPEPKTEPPKP-AAPKKWKPWEDDDPEPEADFTMPAPKK- 1253
QY 1014 TSTMPELNPTSRIAEAMLOTTIRPNTQFNSTKLVENVPKSED-----AGGAEGETP 1063
: : : : :
Db 1254 -----PDTEPDADSLGPKPKDP 1271
RESULT 6
T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N;Alternate names: alpha-NAC protein
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C;Accession: T30826
R;Yotov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A;Title: Differential splicing-in of a proline-rich exon converts alphanAC into a muscle
A;Reference number: Z20889; MUID:96312450; PMID:8698236
A;Accession: T30826
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2187 <YOT>
A;Cross-references: UNIPROT:P70670; EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AAB187
C;Genetics:
A;Gene: Naca
A;Map position: 10
A;Intons: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A;Note: differential splicing converts alphanAC into a tissue-specific DNA-binding activ
C;Keywords: alternative splicing; DNA binding; transcription factor
Query Match 16.9%; Score 996.8; DB 2; Length 2187;
Best Local Similarity 19.2%; Pred. No. 2.4e-12;
Matches 370; Conservative 149; Mismatches 438; Indels 972; Gaps 73;
QY 21 QVSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDFKVCVTAEL----- 68
Db 183 QVPSQTLNLKGT-----PCPPDVTRAFPSSHLENPLASVQPLGM 222
QY 69 SC-----KG-----RCFESF 78
Db 223 SCPQTLNLSNTPVKGVPISSALQSLRLNLKGVSPPARNTAAPSILAPSTSLGCHLPL 282
QY 79 ERGRECDCAOCKYDKCCPDYESFCAEVHNT----- 111
Db 283 LHHSSVDSPIQ-----PPQSGLA-VSNPTSVGHSGIAACPPERCVPALPSRLLA 333
QY 112 -----GPPSSK----- 118
Db 334 VDSGAAPSDDKGSAYTNELCSPPGSSNVAGTSLSPKASLVKPGSNVALQPLVTQVPASQ 393
QY 119 -----APPSGA-----SQIKSTT 133
Db 394 KTGKLEIPVSCIGATHALDNFSAISVAPATHVPPPTSSGLVSKDPASPVTSLVVPAAH 453

QY 134 KRSPK-----PNKKTK----- 146
: : : : :
Db 454 KQFPAPASATLGVVSPFLPATEGLKNLPISALVNVGAPVSPAQAGLPTKRDITLQPLAP 513
QY 147 -----KVIESSEITEK----- 157
Db 514 IALKESPSSQSASLEVLSEDTVTKTTGGPAPVVRPAIAGVATTTSLRADSPPAVIRAD 573
QY 158 --VKDNKNRTKK-----PTPKPPVVDGAGGLDNGDFKVTTPDTSTT 199
: : : : :
Db 574 SCVSNFVSOPLKASVTDPAAPRTAKNTAPSTTSPVLPLASEGCPVASSMALSPQNASV 633
QY 200 QHNKUSTSPKTTTAKPNRPSLPNPS----- 226
: : : : :
Db 634 SETALASPELPKSVFP-PDPPLAEISFNARKVDVSHMESSGSSRQGHDPDASVTAKGT 692
QY 227 -----DTSKETSITVVK----- 238
: : : : :
Db 693 VVCLADSSLOTVSASKSALSGASSPLYPLEVFLPGLAVQGPGLKSLKSLPTPSS 752
QY 239 -----ETTIVETKE----- 246
Db 753 KGAPVSTGAPPPSPKGAPIVPTSESSISKQVPABILPSPKTPBEVTASRLISAVQSPKVD 812
QY 247 -----TTTTNKQSTDKKXTTSKETSQSIETSA-----KDLAPTSKVLAKP- 289
: : : : :
Db 813 PMSDVTPTSPKTSATAVPKDISA--TSLSKSVPAVTSLSPPKAPVAPSNEATIVTEI 870
QY 290 -----TPKAETTTKGPALT----- 304
: : : : :
Db 871 PTSUKNALAAATPKETLATSIPKVTSPSPKTPKSVLKGAPAMTSKKATEIAASKDVSP 930
QY 305 --PKE-----PTPTPKPEASTT-----PKEPTTTIKSAPTTPKP-PAPTTT 344
Db 931 SQPKVEVPLLOHVPPTSPKSPVSDLSGALTSPPKGP-PATLAEITPTPKSPKPAAS 989
QY 345 KSAPTTPK-----EPAPTTKBPAPTTKBPAPTTTKBPAP----- 380
: : : : :
Db 990 KKTATPSPEGTVAVPLEIPPCSKAPKTAAPKESSATSSSKRAPKTAASKVPSKGVTA 1049
QY 381 -----TTTKSAPTTKEPAPTTPKP-PAPTTPKP-----PAPTTT 414
: : : : :
Db 1050 VPLEISLPLKETSUSA--TGEKSASSPKRSPKTAGPKTPPGGVTVAVPPEISLPPKPTP 1107
QY 415 KEPTP-----TTPKP-----PAPTTKEP----- 432
: : : : :
Db 1108 QNATPNESLAASSQKRSKTSVPKTPPGVGTAMPLEIPSAPOKAPKTAVPKQIPTPEDA 1167
QY 433 -----APTTKEPAPTAAPKAPATTP----- 453
: : : : :
Db 1168 VTILAGSPLSPKASKATAAPKEAFATPSGVIAVSGEISPSPKTSKTAAPKENSATLPP 1227
QY 454 -KEPAPTTKEPAPTTTKE-----PSPTTPKS---PAPTTKSAPTTTKEPAPTT 499
: : : : :
Db 1228 KESPKTAAPKE-TPATSEGVAVPSEISPSPTPASKGVFVTLTPKAPNALAE-SPAS 1285
QY 500 TKSAPTT--PKBPS--PTTKPEAPTTKBPAPTTPKKAPATTP-----PKBPAPTTT 547
: : : : :
Db 1286 PKKVEKTAAPBEETSTTSPQKIPKIVAGSEKASATPPSKTKPTKTAVPKETSAPSEGVTAVP 1345
QY 548 KEPAPTTTKBPAPTAAPKE-PAPTTPKETAPTTPKLTPTTPEKLAAPTTPKBPAPTTPEEL 606
: : : : :
Db 1346 LEIPSPRKPAPKTAAPKETPAPS--PEGATTAPVQI--PPSPRKSXKAGSKETPTTP--- 1399
QY 607 APTTPEBPTTTPPEP--APTTPKAAAAP-----NTPKBPAPTTTKE 645
: : : : :
Db 1400 --SPEGVTAAPLEIPISXKTSKAPKETAAPKETAAPKETAAPKETAAPKETAAPKETAAP 1456
QY 646 PAPTTPEKAPATTPKETAAPTTPKETAAPTTPKETAAPTTPKETAAPTTPKETAAPTTP 701
: : : : :
Db 1457 IPPSHKAPKTVDPKQVPLTPSPKADPTTLAE-SPSPSKK-APKTAAPSERVTTVP-- 1512
QY 702 TTSDKAPATTPKGTAPTTKBPAPTTTKEPA-----PTTPKGTAPTTTKEPAPTTT 752

```
Db 1513 ---EKPA-TPQKASGTASKVPVPAETQEVAVSSRETPTVTPAVPPVKPNPSSHKKTKTIE 1568
Qy 753 KKPAPKELAPTTTKGPTSTTSKAPATTPKETATPTPKPAPATTPKAPATTPPTT 812
Db 1569 LKEAPATLPPSPKPKIPSSK---APRTSAP-----KEFPASPKIPVTT 1612
Qy 813 SEVSPTTTTKEPTTHKSDSTPS--LSAEPKPALENSKEPGVPTTKTAPATKPEMT 870
Db 1613 SLAQT-----APPSLQKAPTTPKENLAA---PAVLVSSKSPAAP-ARASASLSP--- 1660
Qy 871 TTAKDKTTERDURTP-----ETTTA-----APKMTETATTEKTTESKITA 913
Db 1661 ATAAPQAPKEATTIPSCKAAATETPTETSTAPSLEGAPKETSTSV-----SKVLM 1713
Qy 914 TTTQVTSITTQDTPFKITLL---KTTTLPKVTITTKTITTEIMNKPEETAKPKDRA 969
Db 1714 SSPPKASSKRASTLPATLPSLKEASVLSSTATSSGK-----DSHISPVSDA 1762
Qy 970 TNSKATTPKPKPTKAPKX-----ETTTA-----APKMTETATTEKTTESKITA 988
Db 1763 CSTGTTT--POASEKLPSKKGTAFTTEMLAAPAPESALAITAPIQKSPGANSASSPKC 1820
Qy 989 --PTSTKPKWPR--VRKPKTTPTRKWTSTMPNLNPTSRIAEAMLOTTPRPNQ----- 1039
Db 1821 PPSKKTGKGLPSAVALAPQTPVEK-----DTSKAIETLLVSPAKGSDCLHSP 1870
Qy 1040 -----TPNSKLVEVNPKSEDAG 1056
Db 1871 KGPVGSQVATPLAAFTSDKVPPEASVASVAPKAPAPASLITLAPSPVAPLPKQPILLESAP 1930
Qy 1057 GAEGTTPMLRPHVMEVTPDMYLRV-----PNQGIIN----- 1094
Db 1931 GSVLESPSKL-----FVPAEDELPLPIPEAVSGGEPFQPLVNMVPAPKPAAGTAP 1982
Qy 1095 -----PML 1097
Db 1983 APSAKQVPL 1991

RESULT 7
T45025
Mucin MUC5B, tracheobronchial (imported) - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T45025
R:Desseyn, J.L.; Guyonnet-Duperrat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A:Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
A:Reference number: 222899; MUID:97166151; PMID:9013550
A:Accession: T45025
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3570 <DSS>
A:Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
A:Experimental source: placenta
C:Genetics:
A:Gene: MUC5B

Query Match 16.9%; Score 992.1; DB 2; Length 3570;
Best Local Similarity 13.1%; Pred. No. 5.9e-12;
Matches 444; Conservative 126; Mismatches 459; Indels 2359; Gaps 103;
Qy 23 SSQDLSSCAGRCG-----EGYS-----RDATC----- 44
Db 172 SAFGTTTCQPCQWTEWDEYKSEQLGGDVESYDKIRAGGHLCCQPKDIEQASFP 231
Qy 45 -----NCD-----YNCQHYMECCPD----- 59
Db 232 NWTLAQGVQKVCVHFGLVCRNWEQGVFKMVCNRYRIVLCCSDDHCRGATTPPTTE 291
Qy 60 ----- 59
```

```
Db 292 LETATTTTQALFSTPQPTSSPQLTRAPPASTTAVPTLSEGLTSPRYTSTLGTATTGGPR 351
Qy 60 ----- 59
Db 352 QSAGSTEPTVPGVATSTLPTRSALPQTTGSLGTWRPSQPTPLAPTTMATSRARPITGTAST 411
Qy 60 -----FKRYCTAELS-----CKGRC----- 74
Db 412 ASKEPLTSLAPLTLSLSELSTQAEETSTPTETMTGSLTNTTTSQGTTRCQPKCEWTFD 471
Qy 75 -----FESPER-----GREC-----DCDAQ----- 89
Db 472 VDPFTSGVAGDMETPENIRAAAGKCMWAPKSIIECRAENYPEVSDQGVLTCSLETGL 531
Qy 90 -CKKYDK-----CCPDYESFC-----AEVHNPTSP----- 114
Db 532 TCKNEQOTGFNMCNPNVNVLCDDY-SHCSTLSTATSTATSPSTPGTWTWTKPTITA 590
Qy 115 -----SKKAPP----- 121
Db 591 TTTASTGSTATASSTOATAGTAPHVSTATTPTVTSKATPFPSSPGTATPALRSTATT 650
Qy 122 -----PSGASOTI----- 129
Db 651 TATSFTALPSSSLGTTWRLSQTTTPMATKSTATSSPTETVHTSNVLTATTATTGATGS 710
Qy 130 ----- 129
Db 711 VATPSTPGTATTKVLTGTTTGTATPSSSPGRATLFPWISTTTTPTTGRSTVTPSSI 770
Qy 130 -----KSTTKRSPKPNKKTKKVIESBEITEKVDKNKN 164
Db 771 PGTHPTVLTNTTAVATGSMATPSSSTQTSGLTPSLTTTATTATGST----- 821
Qy 165 RTKKKP-----TPKPPVV----- 177
Db 822 ---TNFSSPTGTTPIPPVLTATTATPAATSTVTPSSALGTTTTPPVNNTATTGRLS 878
Qy 178 -----DEAGSLDNGDFKVTTPD-----TSTQHNKYSTSPKI----- 210
Db 879 PPSPHVTCVATWATSATSGI-LGTHITEPSTGTSHTPAATTGTTQH-----STPALSSPHS 933
Qy 211 --TTAKPINRPSLPPNSDTSKETSITVNKE-----TTV----- 242
Db 934 SRTESPPSPGTTTTPGHTTATSTTATATATPSKTRTSTLLSQTSAPITVTVMGCEPQC 993
Qy 243 ----- 242
Db 994 AMSEWLDYSYPMFGPSGGPDTYSNIRAGGAVCEQPLGLECRAQAQGVPLRELQVVE 1053
Qy 243 -----ETKETTNTNKQSTDG-----KE 260
Db 1054 CSLDFGLVCRNEQVKFKMVCNRYRIVFCNNGHCPSTPAISSSTATSSSTGTTWILTE 1113
Qy 261 KTSKAKETOSIEKTSKADLAPSKVLAKPTPKAETTKG----- 299
Db 1114 LTTTATTTSTGSTA---TP*STLRATAPPKVLATTATTPTVTSKATPFPSSPGTAT 1169
Qy 300 PAL-----TTP-----KEPTPT-----TPK----- 314
Db 1170 PALRSTATTPTATSTVPIPSLSLGTWTRLSQTTTPTATMTATPSSPETATHTSVLTA 1229
Qy 315 -----EPAST-----TPKEPTPTTI-----KSAPTTPKEP-----APT 343
Db 1230 TATTGATGATVATPSPTPGTATHTTKVPTTTTGTATPSSSPGTALTTPPWISTTTTPT 1289
Qy 344 -----TKSAPTTPK----- 352
Db 1290 RGSVTVPSSIPGTHTATVLTATTTTATGSMATPSSSTQTSGLTPSLTTATTATTGTS 1349
Qy 353 -----EPAP-----TTTKEAPT-----TPKBPATTTTKEAPTTKS-----APT 388
Db 1350 TTNPSSTPGTRPIPPVLTATTATPAATSTVTPSSALGTTTTPPVNTATTGRLSPS 1409
```

QY 389 TP-----KEPAPTTKKEPAPTT-----PKEPAPTTKKE 416
Db 1410 SEHTVRMTAWSATSGTLGTHITEPSTGTSHTPAATGTHQSTDPALSSPHSPSTTESP 1469
QY 417 PTP--TTPKEPAPTTKEPAPTTKE----- 439
Db 1470 PSFGITTPGHTTATSRTTATATPSKTRSTLPLSSPSTSAPIITVVTMGCEPOCAWSEWLD 1529
QY 440 -----PAPTA-----PKKP----- 448
Db 1530 YSYMPGSGGDFDVTYSNIRAAAGAVCEQPLGLECRAQAQGVPLRELQVQVCESLDRGL 1589
QY 449 -----APTTKEPAPTTKE----- 463
Db 1590 VCRNRQVQKFKMCFNYEIRVFCNNGHCPSTPATSSTATPSSTPGTTWILFEQTAATT 1649
QY 464 -----PAPTTTKEPSTTPK----- 478
Db 1650 TATTGSTALPSSTPGTAAPPKVLTSQAATPTATSSKATSSSPRTATLPLVLTSTATKST 1709
QY 479 ----- 478
Db 1710 AFSFTPIPSSTLGTGTSQNRPPHPMATWSTIHPSSTPETHTSTVLTTKATTRATSSM 1769
QY 479 -----EPAPTTTKSAPT----- 490
Db 1770 STPSSTPGTTWILTELTTAAATTAALPHGTPSPSTPGTTWILTEPSTTATVTPGSTAT 1829
QY 491 ----- 490
Db 1830 SSTRATAGTLKVLTSATTTPTVISSRATPSSPGTATALPALRSTATTPTATSVTAIPSS 1889
QY 491 ----- 490
Db 1890 SLGTAWRLSQTTTPATNMSTATPSSTPETVHTSVLTATTTATRTGSAVATPSSPGTAH 1949
QY 491 TTKKEPAPTTT-----KSAPTTPKEP-----SPTT----- 514
Db 1950 TTKVPTTTTGTFTATPSSSPGALTTPVWISTTTPTTRGSVTPSSIPGTHHTATVLT 2009
QY 515 -----TKEPAPTTKEPAP-----TTPK 532
Db 2010 TTTTATGSMATPSSSTQSGTPPSLTATTTATITATGSTINPSSTPGTTPIPVLTAT 2069
QY 533 KPAPT-----TPKEPAPTTTKEPAPTTT----- 555
Db 2070 TPAATSSVTPSSALGTHTPPVNTTATTHGRSLPPSPPHVPTAMTSATSGILGTHI 2129
QY 556 -----KKPAPT-----APKEPAPTT-----KETAPTT 578
Db 2130 TEPSTGTSHTPAATGTTQFSTPALSSPHSPSTTESPPSPGTTTPGHTRGTSTATAT 2189
QY 579 PKK-----LTPTTPEKL----- 590
Db 2190 PSKTRSTLLPSSPSTSAPIITVVTCCEPOCAWSEWLDYSYMPGSGGDFDVTYSNIRAA 2249
QY 591 ----- 590
Db 2250 GGAVCQPLGLECRAQAQGVPLRELQVQVCESLDFGLVCRNRQVQKFKMCFNYEIRVF 2309
QY 591 -----APTTPKEPAPTTPEELAPT-----PEPTP- 616
Db 2310 CNYGHCPSPTATSTATPSSTPGTTWILTKLTTTATTTTESTGSTATPSSSTQGPAGTPH 2369
QY 617 --TTPPEAPTTTPKA----- 629
Db 2370 VSTTATTPVTSSKATPSSPGTATALPALRSTATTPTATSTFAIPSSSLGTTWRLSQT 2429
QY 630 -----AANTPKEPAPT-----TPKEPAPT--TPKEPAPTTPK 660
Db 2430 TTPMATMSTATSSTPETVHTSVLTATTTATTTGATGSAVATPSSTPGTAHTTKVPTTTTG 2489

QY 661 ET-APTTPKGT-----PTT-----LK 676
Db 2490 FTVTPSSPGTARTPPVWISTTTTPTSGSVTPSSIPGTHPTVLTTTTPQPVATGSM 2549
QY 677 EPAPTTTTPKAPKELAPT-----TTKEPTSTTSKAP-----TTPKGTAPT-TP 720
Db 2550 TSPSSSTQSGTPPSLITATITATGSTINPSSTPGTTPIPELTATTAATSSIVTP 2609
QY 721 KEAPAPTTKKEPAPT----- 736
Db 2610 SSALGTTHTPPVNTTATTHGRSLSPSPSHVTAAMTSATSGTLGTHITEPSTGTSHTP 2669
QY 737 KGTAAPT-----LKKEPAP-----TTPKAPKELAPTTTGTSTTSKAPTPKAPT 787
Db 2670 AATTGTTTTPALSSPHSPSTTESPPSP-----GHTTPGHTTATSTTATPSKTRTS 2725
QY 788 T--PKEP--APT----- 796
Db 2726 TLLPSQTSAPITVVTTGCEPOCAWSEWLDYSYMPGSGGDFDVTYSNIRAAAGAVCEQ 2785
QY 797 -----PKKP-----A 801
Db 2786 PLGLECRATAQGVPLGELQVQVCESLDFGLVCRNRQVQKFKMCFNYEIRVFCNNGHC 2845
QY 802 PTTPET-----PPP--TTSVST 817
Db 2846 PSTATSTAMPSSTPGTTWILTELTTATTTASTGSTATPSSTPGTAAPPKVLTSAPT 2905
QY 818 PTTTKEPTTHKSP----- 831
Db 2906 PTATSSKATSSSPRTATLPLVLTSTATKSTATSVTPIPSSTLGTTGTLPEQITPVATM 2965
QY 832 -----DESTPE-----LSAEPKALENGKPGVP-----TTK 860
Db 2966 STHPSSTPETHSTSVLTTKATRAISSSTSPSSSTGTTWILTELTTATTAATGTOPTA 3025
QY 861 TPAATK-----PEMTTAA----- 873
Db 3026 TSPSTPGTTWILTELTTATTTASTGSTATLSSTPGTTWILTEPSTTATVAPPGSTAT 3085
QY 874 -----KDKTTE-----RDLRTTPEPT----- 889
Db 3086 SSTQATAGTHVSTTATTPVTSSKATPSSSPGTATPALRSTATTPTATSTFAIPSSS 3145
QY 890 -----TAAPKMTKETA-----TTTEKTT----- 907
Db 3146 LGTTWRLSQTTTPATNMSTATPSSTPETVHTSVLTATTTATGATGSAVATPSSTPGTAH 3205
QY 908 ESKITATTQ-----VTSITQDTP-----F 929
Db 3206 TTKVPTTTTGTATPSSSPGTALTPPVWISTTTTPTTTTSGSVTPSSIPGTHHTA 3265
QY 930 KITTLKTTTLLA-----PKVTTTKTITTEIMNKPEETAKPKDRATNS 972
Db 3266 RVLTTTITVATGSMATPSSSTQSGTPPSLTATITATGSTINPSSIT----- 3315
QY 973 KATTPKP-----QKTKAPKPTSTKPK-----TMPVRK----- 1003
Db 3316 PGTTPIPVLTSMATTPAATSSKATSSSPRTATLPLVLTSTATKSTATSTFTPIPSSLW 3375
QY 1004 -----PKTTPTPKPM-----TSTMPELNPTSRIAEAMLQTT-----TRNCTPNSKLVE 1047
Db 3376 TWTWVPAQTTTPMTSTMTSIHTSTPETHSTV-----LTATMTATINATPSSSLGT 3431
QY 1048 VNPKE-----DAGGAETGPHMLLR-----H 1070
Db 3432 TRILTELTTTATTAATGSTATLSSTPGTTWILTEPSTIATVMVPTGSTATSTSLGTAH 3491
QY 1071 VFMEVTPDMDYLP-----EVPNGI 1092
Db 3492 T--PKVVTAMATMTATASTVPSSTV 3516

```

RESULT 8
S48478
Glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR019c
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: S48478; B26877; S27281; JG6123
R:Rowley, K.
submitted to the EMBL Data Library, October 1994
A:Reference number: S48478
A:Accession: S48478
A:Molecule type: DNA
A:Residues: 1-1367 <ROW>
A:Cross-references: UNIPROT:P08640; GB:247047; EMBL:Z38061; NID:G603997; PID:G763364; GS
J.Yamashita, I.; Nakamura, M.; Fukui, S.
R.Bacteriol. 169, 2142-2149, 1987
A:Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A:Reference number: A91831; MUID:87194600; PMID:3108330
A:Accession: A26877
A:Molecule type: DNA
A:Residues: 1-242 <YAM>
A:Cross-references: EMBL:M16164; NID:G172522; PIDN:AAA35014.1; PID:G172525
A:Accession: B26877
A:Molecule type: DNA
A:Residues: 762-1331 <YA2>
A:Cross-references: EMBL:M16165; NID:G172523; PIDN:AAA35015.1; PID:G172526
FERS Lett. 239, 179-184, 1988
A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar
A:Reference number: S27281; MUID:89031230; PMID:3141213
A:Accession: S27281
A:Molecule type: DNA
A:Residues: 1-31 <PAR>
A:Cross-references: EMBL:X13857; NID:G4551; PIDN:CAA32069.1; PID:G4552
R.Lambrechts, M.G.; Ianez, E.; Zalcain, M.; Claros, M.G.; Jimenez, A.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A:Title: Muc1, a mucin-like protein that is regulated by Mes10, is critical for pseudohy
A:Reference number: JG6123; MUID:96323237; PMID:8710886
A:Accession: JG6123
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1367 <LAM>
A:Cross-references: GB:U30626; NID:G1304386; PIDN:AAC49609.1; PID:G1304387
C:Genetics:
A:Gene: SGD:MUC1; STA2; NAL5; DEX2; SGD:S0001458
A:Cross-references: M1PS:YIR019c; SGD:S0001458
A:Map position: 9R
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F:5-21/Domain: transmembrane #status predicted <TM1>
F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match      15.7%; Score 923.9; DB 1; Length 1367;
Best Local Similarity 24.6%; Pred. NO. 3.4e-11;
Matches 341; Conservative 113; Mismatches 426; Indels 506; Gaps 69;

QY 5 TLPIVLLLLSVFVYIQVSSODLSCAGRCGEGYSDATCNDYNCQHYMECCPDFKRVK 64
D 129 TFEVY-----ATQDVS-----CQVM--FNFOIQ- 151
QY 65 TAEISCKGRCE-----SFERREC-----DC 86
D 152 -----FEYLOQSAAYASSQWGTTFDLSTGCNNYDNOGHSQTDPPGYMINDC 201
QY 87 DAQC-----KKYDKCCPDY 100
D 202 DNNCGGTKSSITTSSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 259
QY 101 BSFCAEHNPTSPPSKAPPPSGASQTIKTKRSPKPPN-----KKTKKVIREE 153
D 260 -----ATPTTSCTEKPTP-----PTTSCTEKPTPPTPHHDTTPCTKKT----- 300
QY 154 ITEKYDKNKNRTKKKPTPKPVVDAGSLDNGDFKVTPTDSTTQHNKYSTSPKLTIA 213

```

```

301 -----TTSKTCIKKITTVP-----TPSSSTTE-----SSAPVPT- 331
QY 214 KPINPRPSLPNDSITKESLIVNKETT-----VETKETTINKQTS--TQKKTISA 265
D 332 -----PSSSTTSSAPVTSSTTSSAPVTPSSSTTSSAPVTSSTTSSAPVTSSTTSSA 381
QY 266 KETQSIKTSKDLAPTSKVLAKPTKARTTKGPA-----LTPKEPTTTPKEPASTTP 321
D 382 PVTSSSTESS-----APV-----PTPSSSTTSSAPVTSSTTSSAPVTSSTTSSA 428
QY 322 KEPTPTTI-----KSAP-----TTPKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAP 372
D 429 SSSAPVTSSTTSSAPVTSSTTSSAPVTPSSSTTSSAPVTSSTTSSAPVTSSTTSSA 484
QY 373 TTTKEPAPTTKSAPTTKEPAPTTKPKAP-----TTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 422
D 485 VTPSSSTTSSAPVTSSTTSSAPVTPSSSTTSSAPVTSSTTSSAPVTSSTTSSA 536
QY 423 KEPAITTKEPAPTTTKEPAPTTKPKAPTTTKEPAPTTT-----KEPAPTTTKEPAPTTTKEPAP 473
D 537 SSSAPVTSSTTSSAPVTSSTTSSAPVTPSSSTTSSAPVTSSTTSSAPVTSSTTSSA 586
QY 474 PTPKEPAPTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 532
D 587 STTSSAPVTPSSSTTSSAPVTPSSSTTSSAPVTSSTTSSAPVTSSTTSSAPVTSSTTSSA 642
QY 533 KPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 592
D 643 SSSSTSSAPVPT-----PSSSTTSSAPVTPSSSTTSSAPVTSSTTSSAPVTSSTTSSA 694
QY 593 TPEKAPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 643
D 695 VT-----SSTTSSAPV-----PTPSSSTTSSAPVTPSSSTTSSAPVTPSSSTTSSAPVTPSSSTT 746
QY 644 KEPAITTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 702
D 747 SSSAPVTSSTTSSAPVTPSSSTTSSAPVTPSSSTTSSAPVTPSSSTTSSAPVTPSSSTTSSAPVTPSS 800
QY 703 TSDK-----PAPTTKGPATPTTKEPAPTT-----TPKEPAPTTKGPATPTTKEPAPTTKGPATPTT 743
D 801 TSSSVAPVTPSSSTTSSAPVTPSSSTTSSAPVTPSSSTTSSAPVTPSSSTTSSAPVTPSSSTTSSAPV 859
QY 744 LKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 798
D 860 EGSVAPV-----PTPSSSNTSSAPVTPSSSTTSSAPVTPSSSTTSSAPVTPSSSTTSSAPVTPSSST 912
QY 799 KPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 856
D 913 SVSSTTETIVPKTTTSTTTTSTTTTSTTTTSTTTTSTTTTSTTTTSTTTTSTTTTSTTTTSTTTTSTTT 967
QY 857 PITKTPAATKPEMT-----TAKDKTERDLRTTET-TTAAKPKMKETATTTTEKTE 908
D 968 PTTTTSVTSTTTTSTTTTSTTTTSTTTTSTTTTSTTTTSTTTTSTTTTSTTTTSTTTTSTTTTSTTT 1027
QY 909 SKITATTTQVIST--TTQDTTTPFK-----ITLKTTLA--PKVTT-----TK 947
D 1028 SPPTPTVTVSVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTVTV 1087
QY 948 KITTTETI---MNKPEET---AKPK----- 966
D 1088 TTTTTCSTGNSAGETTSGCSPKVTTTTPVCSGTGTYTTEATTLVTTAVTTTVTTTE 1147
QY 967 -DRATNS--KAT-----PKQPKTPAKPKP----- 989
D 1148 SSGTNSAGKTITGTTKSVPTTVTTLAPSAVPTPATNAVPTTITTECSAATNAAGET 1207
QY 990 TSTKPKKTM-----PRVRKPKTPTPRKMTSTMPELN-----P 1022
D 1208 TSVCSAKTIVSSASAGENTAPSAVPTTPTTPTTITTESSVGTNSAGETTGTGTTKTSIP 1267
QY 1023 TSRIAEAM-----LQTTTRPNOTPNSKLVNPKSEDAGAGETPHMLLRPHV 1071

```

Db 1268 TTYITTLIPSGNAGNYETVATATNP-----ISIKTTSQIATATASASS-----VAPVW 1315

Qy 1072 FMPEVT 1077

Db 1316 TSPSLT 1321

RESULT 9

T34513

hypothetical protein ZK783.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004

C:Accession: T34513

R:Favallo, A.; Vaudin, M.

submitted to the EMBL Data Library, August 1994

A:Description: The sequence of C. elegans cosmid ZK783.

A:Reference number: Z21536

A:Accession: T34513

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 13507 <FAV>

A:Cross-references: UNIPROT:Q23587; EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783

A:Experimental source: strain Bristol N2; clone ZK783

C:Genetics:

A:Gene: CRSP:ZK783.1

A:Map position: 3

A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2; 3504/1

Query Match 15.5%; Score 910.5; DB 2; Length 3507;

Best Local Similarity 13.3%; Pred. No. 2.3e-10;

Matches 405; Conservative 183; Mismatches 448; Indels 2005; Gaps 94;

Qy 7 PIVLLLLSVFVIQVSS-----QDL-----SSC 30

Db 32 PTFVVPDSTVICQHSDDPTDLIHNMSSLDGKQDCFNPAHHDVEVPYCEHKQSTC 91

Qy 31 AGR-----CGEGYSRDA-----TCNC-----46

Db 92 SGRGACLYDCAKPCQCYDSFGSACELQDNLEHPCHMAQCONTLGSEYECRLPGY 151

Qy 47 ---DYNCQHWEC-----CPDFKRV-----63

Db 152 EGNHECTDIDECSDLTSCRPHSKCINLPGYNYCQTGFTPKNQSGLDKCADINE 211

Qy 64 ---CTAELCK-----GRCFESFERGRE-----83

Db 212 CETGAHNCDADEICENSIGSFKVNKCPGYELIDGKCEDVNECGSEKHLKCDVRADCVN 271

Qy 84 ---TGGECECEBEGFDGKNCQKQKSHSVKLPKSKCRKNSAICDRHASCHIVLDICDKT 331

Db 87 ---DAOCK-----92

Db 332 GYTDGITCHINECDKADTPCSGGRCNLNDGGYVCCNKQDDATCIKQGAFCSGGG 391

Qy 93 ---YDKC-----CPDYESFCAEV-----HNPTS 112

Db 392 DNAICSNATCACIDFRGDPHKKVCVDINECVENDSVCGVGDRCVNLFGGFKCCQHGSTE 451

Qy 113 PPSKKA-----PPPGASOTIKSTTKRSPK-----138

Db 452 ABCTDQAFSSDSTISSHGADFTTGGQIIEGSGSIQTSSGSLTV--TRGLIPKQVELT 509

Qy 139 ---PPN-----141

Db 510 TSGRLACTSYCFPNSECVGVCYGVGNALVGCEDIDECITEICNEANEWCNVNLIG 569

Qy 142 ---KKTKKVISEEE-----153

Db 570 GFVCCNPNTAHDICIDFLTKVKIYAYMIIFLLKGLLEITKEKGLHVGNEEDTVVATR 629

Qy 154 -----ITEKVDKNKURTKK-----168

Db 630 SNHSTSDQLITQVQQRNFSTGOIILTRCKVSSGEAVTOTTDADEFGLEISAADLAGSG 689

Qy 169 ---KPTPKPPVDEAGSGL-----DNGD-----188

Db 690 SGITLPTTLEPKIE--GSGKASGVWTEDEGEDEDLMEEGSGSWTTINGITGSPR 747

Qy 189 ---FKVTT-----PDTSTT-----199

Db 748 SEGIRVRIITLGDGEPEFATKPGISAPDKTCEGSKTESDGEKLTVEKDGEAQSSES 807

Qy 200 -----QHNKVTSPKI-----TTA 213

Db 808 SATSSGKKSEATSGSSSSAKSGTSEASGSSGASSSSSGSGSGSGSSSVTSEGSFETS 867

Qy 200 -----QHNKVTSPKI-----TTA 213

Db 868 SSGSVGEATGTVDGSESGKPSKSTEEKLPFTKNGEKSPISGSDITTKESSEETSR 927

Qy 214 KPINRPRSLPPNS-----DTSKETSLTVNK-----238

Db 928 KPIEGSDSLTSGSGSGWFEFGKSGHFESGSKVSTGKGTQSGAGSGSGPKVPKPG 987

Qy 239 ---ETTVEKETTIT-----NKQSTDCGKETTSAKETQSI 271

Db 988 APEITTDGEESSTGDKGGKPADKSNKVPKTDGDKNPDIITDGEDSTSETSGGEQG 1047

Qy 272 EKTSAKDLAP-----TSKVLAKPT-----290

Db 1048 PKGSKGPPGDKGSEVKKPTSEVDGPNLGGKSNVPLKPTDLPPEGSGILTSSGG 1107

Qy 291 ---PK-----AETTKGPAITPKPTPT--TPKE-----PA 317

Db 1108 KNSTPEHGTLERLPPKTEDKSETPOLGLEISAGKKPEPEDGTSKEVGLLEIWESTTPG 1167

Qy 318 STT-----320

Db 1168 STTLDSDSVGLIEISGDLTKATKPHVIEGSGTGDEIATTRDVSKSTKKPRVEVDGG 1227

Qy 321 ---PKPEPTTIKSA-----PTT-----PKPE-----339

Db 1228 DNGETSGVDGKPTTAPTPSSASBSTSRIPTTSEASPEGSGSAGVPEGSGSESSTS 1287

Qy 340 ---APITTKSAPITPKPEAPT-----TKPEAPITPKPEAPT-----374

Db 1288 APDGVSPSTTAPETVPTTSASSTPDVAEESGIPSTSKPTAEPLTT--APSTEVTS 1343

Qy 375 -----TKPEAPITTKSAPITPKPEAPT-----KKPAPITPKPEAPT-----413

Db 1344 ESGGTEESTLPTTEGSGESTTSSAPT--EPATVLPQNRNEKPEPTKDTALPTTTTGA 1401

Qy 414 -----PKE-----416

Db 1402 PQANDSSVENTKCTSSDEGLDALCERRTGVCRCPEGPEGAPPKSCVDVDECATGDHNC 1461

Qy 417 -----PT-----418

Db 1462 HESARCONVGVYACFTGFRKADGSCQDIDECTEHNSTCCGANAKVKNKPGYSCCEC 1521

Qy 419 -----PTTPKPEAPITK-----430

Db 1522 ENGLDGVQCVPTT--KKPCDSTQSSKSHCESNMSCVDTVDGSECKECCGKYKSGK 1580

Qy 431 -----EPAP-----434

Db 1581 VCEINECVAEKAPCSLNNANVMNMGTFSCCKQGYRGDGMCTDINECDERHPCHPAE 1640

Qy 435 -----TPKE-----439

Db 1641 CTNLEGSFKCECHSGFEGDGKIKTNPLERSCEDVEKFCGRVDHVSCLSVRIYNGSLSV 1700

Qy 440 -----PA-----441

Db 459 -----MPTPHSPAD-----DYVPTTPVPGKSPATSPSPQVQPPAATPP 500
Qy 220 PSLPPNSDTSKTSLSLVNKETTVETKETTINKQISTDCKEKTSAKETQSIETKSADL 279
Db 501 PSL-----
Qy 280 APTSKVLAKTPKAEITTKGPALTTKPTPTT-----
Db 507 SPQAPAVGPPPPVKTSPAPAGISPPPPVSVVSPPPVKSPPPPAPVGSPPPPPKSP 566
Qy 313 -PKPASTTP--KETPTTIKSAITP-KEPAPTTKSAITP--KEPAPTTKEPAT 365
Db 567 PPAPVASPPPPVKSPPPPPLVASPPPPVKSPPPPAPVASPPPPVKSPPPPPTFVASPPPP 626
Qy 366 TPKEPATTTKEPATTTKSAITP-KEPAPTTKSAITP-KEPAPTTKSAITP-----PKE-- 416
Db 627 APVASSPPPKSPPPPTPVSSPPPPPKSPPPPPPKSPPPPPPKSPPPPPPKSP 686
Qy 417 -PTPTTKPAPTTKAPATTTKAPATTKAPATTKAPATTKAPATTKAPATTKAP 472
Db 687 LPPPTLIPPPPPKPTPTPTSPK-PSGPKSP--PKPVSPPPTKSPPP-----P 738
Qy 473 SPPTTKPAPTTKSAITP-KEPAPTTKSAITP-KEPAPTTKSAITP-KEPAPTTK 532
Db 739 APVS--SPPTTPVSSPPALAPVSSPPSVKSP--PPAPLSPPPAPQVKSPPPPVQS 792
Qy 533 KPAPTTKAPATTKAPATTKAPATTKAPATTKAPATTKAPATTKAPATTKAP 584
Db 793 SP--PPAPKSPPLAPVSSPPQVKTSPPPAP-SSPPLAPKSPPPHVVSPPPPVKSPP 851
Qy 585 TTEKAPATTKAPATTKAPATTKAPATTKAPATTKAPATTKAPATTKAPATTK 638
Db 852 PAVSSPPPTKAPASPPAHVSSPPVVKSTP--PAPTIVSP-PSEPKSPPPPTPVSLP 908
Qy 639 APPTKAPATTKAPATTKAPATTKAPATTKAPATTKAPATTKAPATTKAP 687
Db 909 PPIVKSSPPAMVSSP-PMTFKSSPPVSVSSPPPTVKSSPPAPVSSPPATPKSPPPA 967
Qy 688 PKELATTKE---PTSTSDKAPATTKGATTP-KEPAPTTKAPATTKGATPT 743
Db 968 PVNLPPPEVKSSPPPTFVSSPPA---PKSSPPAPVSSPPPPPKSPPPPPAPVSSPPPP 1024
Qy 744 LKAPATTKPKKAPKELATTKGPTSTSDKAPATTKGPTSTSDKAPATTKGPT 797
Db 1025 VKSPP-----PAPVSSPPPVK-----SPPPAPVSSPPPPVKSPPPPAPVSSPP 1071
Qy 798 --KKAPATTP-ETPPTTSEVSTPTTKPTTHKSPDSSTPBLSAEPTKALENSPK 854
Db 1072 PVKSSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPPKSPPPPPAPVSSPPPPA--VKPP 1125
Qy 855 GVPTTKTAPATKEMITTAADKTTEDLRTPTTTTAAKMTKATTTTEKTTESKITAT 914
Db 1126 SLP-PPAPVSSPPVTPAPPKKEOSL-----
Qy 915 TTQVSTTTQDTPFKITTLKTTTLAPKVTTTKTITTTTINMKPEETAKPD----- 967
Db 1153 -----PPAESQPPSPFNIDILP 1170
Qy 968 -RATNSKATTPKQ 980
Db 1171 PIVANKYASPPPPQ 1184

RESULT 11

T29757
protein UNC-89 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
C/Accession: T29757
R/Du, Z.; Le, T.T.; Wilson, R.
submitted to the EMBL Data Library, May 1997
A/Description: The sequence of C. elegans cosmid C09D1.
A/Reference number: Z20679

A/Accession: T29757
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-6642 <DUZ>
A/Cross-references: EMBL:AF003131; PIDN:AAB54132.1; GSPDB:GN00019; CESP:unc-89
A/Experimental source: strain Bristol N2; clone C09D1
C/Genetics:
A/Map position: 1
A/Gene: CESP:unc-89
A/Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1;
/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Query Match 14.8%; Score 872.9; DB 2; Length 6642;
Best Local Similarity 13.9%; Pred. No. 3e-09;
Matches 398; Conservative 144; Mismatches 382; Indels 1941; Gaps 95;
Qy 34 CGEGVSRDATCNC-----DYNCOH----- 52
Db 659 CKEGDIVDFCEVGEVPEPELVWLDQPLRPSHDFLQDQGTAKLEIRDAQPDPTGVY 718
Qy 53 -----YMECCPD-----PKRVCTAB-- 67
Db 719 TVKIQNEFGSIESKAELFVQADPDKNHVAPEFOATIEYVEC--DEGBEVAFKSVITDPN 776
Qy 68 -----LSCKGRCFESFERGREDC 86
Db 777 PEIWFINGKPLSSEKVKFTSEGGICILITIKOVTRHFDGMVTCGS-----NLGSASC 831
Qy 87 DAQCKKYDKCCPDYESFCAEVHNPTSPSSKKA-----PPPS-- 123
Db 832 DGRLL-----VRVPPAPPTFNKPLEDKTVQEKSTVFEVDVSGWPEPTLT 876
Qy 124 -----GAS----- 126
Db 877 FTLCGKELNGEGVEIVGHDFYRISIPNTSMKDHPGEIVAKAQNHEGHTAESARLTVE 936
Qy 127 -----QTLK-----STTKRSPKP----- 139
Db 937 QEEBESRSAPFLKDIEDQVKTGEFAVFETTVGNPNPEVTWINGHMDQSGPVKIE 996
Qy 140 -----PNK-KTKKVIE-SEETEK 157
Db 997 AHNHDHKLITDSAOVAGTVLCRAENAVGRFETKARLVVLAPEKOKKPKFVEILVDKET 1056
Qy 158 VKDNKNRTKKPKPKPV-----VDAG----- 181
Db 1057 VDNVTVFVRVEGPKPTVTWYLGEBELKQSDRVEIREFDGSIKISIKNIKIEDAGEIRA 1116
Qy 182 -----SGLDNGDFKVT-----TPDTST-----TQHNKVS----- 192
Db 1117 VATNSEGSDETKAKLITVQKPPAPEFDLRPVSLTVEKGEAVFSAHAFGIPLTVWSVN 1176
Qy 193 -----TPDTST-----TQHNKVS----- 205
Db 1177 GRKVRDQEGARVTRDSTVDGASILTDTATYVSEVNHLTISVVAENTLGAETGAQLT 1236
Qy 206 -----TSPKITTAKPINRPSLPNPPNSDTSKETSILT 235
Db 1237 IEPKESVVEKQDLSSSEVQKEIAQVKEASPEATIT-----ITMETSILT 1282
Qy 236 VNKETTVETKTTT----- 250
Db 1283 STKTTTMTSTTEVTSTVGTVVETKESSESATTVIGSGGVTEGSISVSVKSTDS 1342
Qy 251 -----NKQTSITDGEKTTSAKETQSIETKSADL 279
Db 1343 QTDVREGTPKRVSAEELPEKVIDSDRKKKSPSPKKEK-----SPEKTEKPA 1394
Qy 280 APTSKVLAKTPKAEITTKGPALTTKPTPTT-----KEPASTTPKEPT 327
Db 1395 SPTKKT-----GEEVKSPEKSPASPTKKEKSPAABEVKSPPTKKEKSPSPPT 1441
Qy 328 TIKSAPTT-----KEPAPTTKSAITPKEPATTTKEPATTT-----KEPA 371

Db 1442 KKEKSSPTTKTGDVEKSP-----PKSP-----TKKESPEKPEDVKSPVKKEKS 1489
QY 372 PTTT-----KEPAPTTKGAFTT----- 390
Db 1490 PDAATNIVESSSETTIEKTTTMTTHESBSRTSVKKEKTEPKVDEKPSPTKKDKSP 1549
QY 391 -----KEPAP--TTPKKAPATTPKGPAPTTPEKPTTPKPE--PAPTTPKEPAP-- 434
Db 1550 EKSITBEIKSPVKKEKSPKVEKAPSPATPKKESPEKSPATPKKSENEVKSPTKKESPE 1609
QY 435 -----TTPKGPAP--TAPKKAPATTPKGPAPTTTPEKAPATTPKPSPTTPKGPAPTTTK 486
Db 1610 KSVVEELKSPKESPEKADDPKSPKKE-----KSPKESATEDVKSPTKKESPEKVE 1663
QY 487 SAPTTTKEPAPTTTKSAPT-----TPKPSPTTKE--PAPTTPKEPAP----- 528
Db 1664 EKPT-----SPTKESPTTKTDDVKSPTKKESPTQVEEKPEASPTKKESPEKSVVEE 1718
QY 529 -TTPKKAPATPKPE--PAPTTPKEPAP-----TTPKK-----PAPTAPKSPAPTTPKET 574
Db 1719 VKSPEKSPKAEERKSPKTKKESPEKSAABEVKSPTKKESPEKSAABEVKSPTKKES 1778
QY 575 APT-----TPKKL-----TPTTPEKLAPTTPEKAPATTPPELAPTTPEEP 614
Db 1779 SPVQADDEVKSPTKKESPEKSPKAPATPKKE-----TPEKSA--ABELKSPTKKEK 1831
QY 615 TPTTP-----EBAPTTPKAAANTPKPEPAPTTPEKAPATTPKPEPAP-- 656
Db 1832 SPSSPTTKTGBDSKESPEKPEKSPKSPKSPKSPKSPKSPKSPKSPKSPKSPKSPKSPK 1886
QY 657 ----- 656
Db 1887 LTRDLKLTQVNTKDLAHPVVVVEHATECKWFLDGGKEITTAQGVTVSKDDQPEFRCSIDTT 1946
QY 657 -----TTPKET----- 662
Db 1947 MFGSGTVSVVAGSAGSVETKELVLETPEKTEKPEFTDKLRDMEVTKGDTQVMDVIAL 2006
QY 663 -----A 663
Db 2007 HSPLKYQNGNLLBEGKNGVTIKNEENKSSLIIPNAQDSKGIIVASNEVGSSESAQL 2066
QY 664 PTPKGTAPTTLKEPAPTTPKKAPKAPKELAPTTKEPTSTTSKDPAPTT----- 711
Db 2067 TVNPSSTPIVVDGPKSVTIKE-----TETAEFKATISGFAPTVKWTINEKIVEES 2118
QY 712 -----PKGTAPTTPK-----EPAPTTPK----- 729
Db 2119 RTITTIKEDVYTLKISNAKIBQGTGVKVTAGNSAGQDSKQADLKVEPNVAPKPKSQLT 2178
QY 730 -----EPAPTTP-----KGTAPTTLKEPAP-- 749
Db 2179 DKVADEGEPLRWNLDELDPSPGTEVSVLLNGQLTKSDTVQVVDHGDGTGTVHTIAEAKPE 2238
QY 750 -----TTPK-----KPA 757
Db 2239 MSGTLTAKANAAGCE--SAKVTVNGNKKPEFVQAPQNHETTLSESVKFAIVTGKPMF 2298
QY 758 K----- 758
Db 2299 NVTWYLNKKLIQSEEVKVKYVHETGKT--SIRIQPLMEHNGTIRVAENVSGKVAQAL 2358
QY 759 -----ELAPTT-----KGPTSTTSK 775
Db 2359 KVDKXTEVPKFTNTNDDRVKQKEDGVKFTANVEGYPEPSVAWTLNGEVPKSHPNITVTDK 2418
QY 776 P-----APTTPEK-----APTTPK----- 790
Db 2419 DGEHTIEISAVTPEQAGELSCAATNPVGSKKRDVQLAVKKVGDAPTTAKNLEDRLITEGE 2478
QY 791 ----- 790

Db 2479 LTLMDAKLNTVKKPKKITWLKOGVEITSDGHYKIVBEEDGSLKLSLQTKLEDKGRITIK 2538
QY 791 -----EPAPTT----- 796
Db 2539 AESEFVAGCSASLVGVKGRPAQFQSDIAPINLTGDTLECKLLITGDPFFVKWVI 2598
QY 797 -----PKK-----PAP 803
Db 2599 GTQLVCATEDEISNANGVVTMKIHGVTADMTGKIKCVAYNKAGEVSTEGPKVWAPIV 2658
QY 804 TPET-----PPPTT-----SEVSTPTTTKE 823
Db 2659 EFETSLCDATCREGDTLKLRPAVLGPEPVPVSWYVNGKLEBSQNIKIHSEKGTIVTIIK 2718
QY 824 PTT-----IHK----- 829
Db 2719 DIICDYGQVVCEAINIYKATSEATLLVLRGEPDPFLEWLSNVRAITGTVVHKVFT 2778
QY 830 -----SPD----- 832
Db 2779 GDPKPSLTWINKKEILNSDLYTIVTDDKTSILTINSFNPDPVHVGEIICKAENDAGEVSC 2838
QY 833 -----ESTPELSAE-----PTPKALENSPK----- 852
Db 2839 TANNITVTSDFSESESAQAEBFVGDDLTEDESLREEMHRTPTPW--APKFITKMD 2895
QY 853 -----EPGVPTTK----- 860
Db 2896 TKAKGHSVFCVCPDPTKGVCCKWLKOGKEILARIQVTRTGPGEHITBELVDNVT 2955
QY 861 -----TPAAT-----KPEMTTAKDKTTERDLRT----- 885
Db 2956 PEDAGKYTCIVENTAGDKICEATLVIESLEKSEKKAPEFIVALQDKTKTSEKVLBEC 3015
QY 886 -----PETTHAAPQVKTETATTTEKT--TESKITATTTQ-----VTSTTQDTTPEKIT 933
Db 3016 KVIGBPKPKVSWLHDNVSEKPNSEKTIITQESITVESVEGVERVITSE----- 3066
QY 934 L-----TPTTLPKVTITTKT-----ITTEIMNKP----- 959
Db 3067 LSHQGYTCIAENTEGTSKTEAPLTVQGEAPVFKELQNKELSIGELKVLSCSVKGSQP 3126
QY 960 -----ETAPKPKOR-----ATNSKAT 976
Db 3127 HVDFYGFSETKVETKITSSRIAEHDQTNTHWRMVISQITKEDIVSYKAIATNSIGTA 3186
QY 977 PKQKPT--KAP-----KKPTSTK-----KP----- 995
Db 3187 TSTSKITTKVEAPVFPQGLKTSVKEKEIKMBVKVGSAPDVENFKODKPVSEDNHEM 3246
QY 996 KTMPR-----VRPKPT-----PT-PRKMTSTM 1017
Db 3247 KKNPETGVFTLVKQAATTDAGKYTAKASNPAGTAEASAEVTSQLEKPTFVRELVTTE 3306
QY 1018 PELNPTSRITAEAMLOTTT-----PQTPNSKL----- 1046
Db 3307 VKINET-----ATLSVTVTKGVDPDPVSWLKDQPVQDTSVHVIKVEGSGSYITIKDAR 3361
QY 1047 -----EYNPKSESDAGAGETPHMLLR---PHVFMPEVTP 1078
Db 3362 LEDSGKYACRATNP---AGEAKTEANFAVVKLVPPPEVEKLSP 3402

RESULT 12

T25697
hypothesized protein F16f9.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T25697
R;Fulton, B.
submitted to the EMBL Data Library, August 1996
A;Description: The sequence of C. elegans cosmid F16f9.
A;Reference number: Z20071

Db 206 TAASQAPKATVEVKFAAAVAKAAYATAAAAPKATAEAKPAPVTSPTTICSAEAKP 265
 QY 309 ----TPTTPKEPASTTPEKPTTTIKSAPTTPKEPAPTTPKSAPTTTPKEPAPTTPK 360
 Db 266 LTAASPTASKATAEAKPVATASLMAKVTAEAKPAP-----SPSVPK--ATTDTKAVTA 318
 QY 361 ----EPAPTTPKEPAPTTPKAPAPTTPKSAPTTTPKSAPTTTPKSAPTTTPK 394
 Db 319 TAPKAGPDVKPAPAVCAEAKPAPPPPPQQLPKAAAAAAAPTGTTELKAPATAPPHGFRANSH 378
 QY 395 ----PTTPKAPAPTTPKAPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 448
 Db 379 TVTVTPPNVRAAATVP--TAGAVPKASTGTP-----AAAPQGPV-----KA 421
 QY 449 APTTPKEPAPTTPKAPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 501
 Db 422 APVTPPSPOQAVPR-----AATAAAAPVTPQOP---VTKAAATTTNATPPPPQIPKAAATTTT 474
 QY 502 SAPTPKPSPT--TTKAPAPTTPKAP-----APTTPKAPAPTTPKAPTTPKAPTTPK 547
 Db 475 ATPVTPQQIPKAGTDAAPPPAPKAPSDGRAATPGVPAATDPKPPPTTQSVPSAVTE 534
 QY 548 KEPAPTTPKKA--PTAP-----KEPAPTTPK-----ETAPTTP 579
 Db 535 PKQPRAAPPSPNEATPAVPSPNLKSPLPTPKPVPLMALTPQPVTAQMVTLAATKP 594
 QY 580 KLTPTTPEKLAPTTPKAPTTPBELA-----PTTP-----EPTPTTP 619
 Db 595 SPIVPKASPK-ALMTTPPPPPGLPRALAAKLLGLPSFVASAMHAKVTPRPLPASPVFM 653
 QY 620 -EAPAPTTPKAA----- 630
 Db 654 AASPASLGDAARVALATNAASPGAKPEAGNGTLMAPMGAANTQMAPIGAGAAQATP 713
 QY 631 ----APNTPKP--APTTPK 644
 Db 714 MGAATHVSPMGAGGATQMSPTGAANTHMSPIGAGGATQMSPMGAANTQMSPMGATTTQM 773
 QY 645 EP---APTTPKAP-----APTTPKETAPTTPKGT---AP 672
 Db 774 SPWGAATTPQSPMGAATQVATAGNTQVSPMGAATTPQTSVGAATTPQSPMGAA 833
 QY 673 TTLKEP--APTTPKAPKELAPTTPKAP--TSTTSDKAP-----TTPKGAPTTPK 721
 Db 834 TTLMSPMGAATTPQ--PSPMGAATTPQPMGAATNTTQPPMAASTPQSTPMGAATTTQS 890
 QY 722 EP--APTTPKAPTTPKGTAPTTLKAPTTPKAPKELAPTTPKGTSTTSKAPATT 780
 Db 891 PPMGATTTQSP-----PMGA-----STQAP-----PTVAGSPT-----PPPIP 925
 QY 781 PKETAPTTP--KEPAPTTPKAPTTPPETTPPTTSEVSTPTTTPKPTTI---HKSPDE 833
 Db 926 PSPTAQTSPQMSKSPDPDPKAPSAAGTSPAAHVANASPGVAVSPAPIGVTEASPSA 985
 QY 834 STPELSAETPKALENSKEPQVPTTKTAPATKPEMTTAKDKITRDLRTPPTTTAAP 893
 Db 986 DGAELS--PGPTAATDGK-----ASPAAT--ADVTEAATDVTA-----AATAVPAEAP 1031
 QY 894 KMTKETATTTETKTESKITATTQVSTTQTTQTTTTPFKITTLTKTTLAPKVTITTKTITTT 953
 Db 1032 TKAKRSS 1063
 QY 954 EIMNKPEETAOKDRAATNSKATTPKQKPTKAPKPTSTKPKPMVRKPKTTPTPPKM 1013
 Db 1064 -----SSSSSSSESNPASPA-----PAVGDGQQQWTPGAA 1092
 QY 1014 TSTMPENFTSRIAGMLQTTTRPQNTFNSKLVEVNPKSSEDAGGAE 1061
 Db 1093 QSVPP-----VTEAAVO-----EAAAAAAGAGERE 1119

cell proliferation antigen Ki-67, short form - human
 C/Species: Homo sapiens (man)
 C/Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 15-Mar-2004
 C/Accession: B48666
 R/Schluter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerdes
 J. Cell Biol. 123, 513-522, 1993
 A/Title: the cell proliferation-associated antigen of antibody Ki-67: a very large, ubiqui-
 ins.
 A/Reference number: A48666; MUID:94043435; PMID:8227122
 A/Accession: B48666
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-2897 <SCH>
 A/Cross-references: EMBL:X65551
 C/Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
 F;29-91/Domain: Kinase interaction domain homology <KIH>
 Query Match 14.6%; Score 858.5; DB 2; Length 2897;
 Best Local Similarity 15.9%; Pred. No. 1.8e-09;
 Matches 418; Conservative 132; Mismatches 450; Indels 1629; Gaps 100;
 QY 4 KTLPIVLLLLLVFVVIQQVSSQDLSSCAGRCGEGYSDATC----- 44
 Db 241 KTAP-----ASSSKSQTEVPKRGGE---RVATCLQKRVISRSOHDILQMT 284
 QY 45 -----N 45
 Db 285 SKRSGASEANLIVAKSWADVVLGAKQTQTVKIHGPORSMNKRQRRPATPKPVGEVH 344
 QY 46 CDYNCOHYMECCP-----DFKR----- 62
 Db 345 SQPSTGHANSPTIILGKAHTEKVHPARPYPVLANFISNQXMDFKEDLSGAEAFKTFV 404
 QY 63 -----VCTAELS-----CKGRCPESFERGECDDAQCKYVDCPPDYEFCAEV--- 107
 Db 405 KEOPQTSTCHIAISENENLLKQFGTDSGEE-----PLLPTSEFGNVFVS 453
 QY 108 -HNPTSPPSK-KAPPP-----SGASOTIKST 132
 Db 454 AQNAAQPSDKCSAPPLRRQCIRENGVAKTPRNTYKMTSLTSTDTETEPSTV-ST 512
 QY 133 TRKS-----PKPPNKKT----- 145
 Db 513 VNRSGSTEFRTQKLPVESKSEETNTEIVECILKKGQKATLLQORREGEKIEPPT 572
 QY 146 -KVIIESEITEKVDKNNKNT-KKPKTPKPPVVDAGSLDNGDFKVTPTD----- 196
 Db 573 YKENILKENDKCKAMKRSRTWGQKCAPMSDLTD-----LKSPLDTMLMKDTA 621
 QY 197 -----STQHNVKSTPKITTAKPINRPSLPP-----NSDTS-----KETS 233
 Db 622 RGQNLQTDHAKAPKSEKQITK--MPCQSLQPEPINTPTHTKQOLKASLGKGVKEEL 679
 QY 234 LTVNKETTVETKETTNTKOTSTDGKE-----KTTSAKETQ 269
 Db 680 LAVGKTRT-SGETTTHREPAGDGKSIITFKESPKQILDPAARVTGMKKWRTPKEEAQ 738
 QY 270 SIKTSKADLAPTSKVLAKPTPKAE-----TTTTK----- 298
 Db 739 SLE-----DLAGFKELFQTPGSEESMTDEKTTKIACKSPPEVDPTPTSTQWPKRSLR 793
 QY 299 -----GPALTTPKEP-----TPT----- 311
 Db 794 KADVEEFIALRKLTPSAGKAMLTTPAGGDEKDKAFMGTPVQKLDLAGTLPGSKEQLQ 853
 QY 312 TPKEPAST-----TPKEP-----TPTTIKSAPT-- 334
 Db 854 TPKEQALEDLAGFKELFQTPGHTBELVAAGKTTKIPCDSPQSDPVDPTPTSTKORPKES 913
 QY 335 -----TPKEPAP-----TTT 344
 Db 914 IRKADVEGELLACRNLMPSAGKAMHTPK---PSVGEKDIIFVGTVPQKLDLTENLTGS 970

QY 345 KSAPPTPKKPAPT-----TTKEPAPTTPKEPAPTTPKKEPA 379
Db 971 KRPPQTPKEAQALEDLTGFKELFQTPGHTTEAVAAGTKTKMPCSSPPESADT----- 1024
QY 380 PTTTKSAPTPPK----- 392
Db 1025 PTSTRQPKTPLEKRDVQKELSGALKKLTQTSGETTHDQVPGGDKSINAPRETAKQKLD 1084
QY 393 -PAPTPPKKPAPTTPKEPA-----PTTPK-----BPTPT- 420
Db 1085 PAASVTGSRGHPKT-KEKAQPLEDLAGWKELFQTPVCTDKPTTHKTKIACRSQDPVD 1143
QY 421 TPKEPAPTTKE-----PAPT----- 435
Db 1144 TPTSSRQSKSLRKVDVEEFALRKRTPSAGKAMHTPKPAVSGEKNIYAFMGTPVQKL 1203
QY 436 -----TPKEPA-----PTA 444
Db 1204 DLTENLTGSKRLQTPKEKAQALEDLAGFKELFQTRGHTTEBMTNDTKAVACKSQDPL 1263
QY 445 PKKPA-----PTTPKEPAPT----- 459
Db 1264 DRNPASSKRLKTSIGKVGKBEALLAVGKLTQTSGETTHHTPTGDKGSKMAFMESPKQ 1323
QY 460 -----TPKEPA-----PTTPKEPAPTTPK-----E 479
Db 1324 ILDSAASLTGSKRLQTPKKEVPEDLAGFIELFQTPSHKTE-SMTNEKITTQVSRASQ 1382
QY 480 P-----APTTPKSA----- 489
Db 1383 PDLVDTPTSSKQPKESLRKADTEEFALFRKQTPSAGKAMHTPKPAVSGEKNINTFLT 1442
QY 490 -----TTTPKEPA-----PT-----TTK-----SA 503
Db 1443 PVOKLDQGNLPGSNRRLOTRKEKAQALBELTGFRLEFQTPCTDNPDADEKTTKILCKS 1502
QY 504 PTPPKPSPTTK-----PA----- 519
Db 1503 PQSDPADTPTNKQPKRSLKADVEEFALFRKLTSPSAGKAMHTPKPAVSGEKNINTFV 1562
QY 520 -----PTTPKEPAPT-----TP----- 531
Db 1563 GTPVEKLDLLGNLPGSKRRPQTPKEKAQALEDLAGFKELFQTPGHTTEBMTDDKITEVSC 1622
QY 532 KKPAPTTPKEPA----- 543
Db 1623 KSPQDPVKTPTSSKORLISLGVGVKEVLPVVGKLTQTSKTTQTHRETADGKSIA 1682
QY 544 -----PTTPKEPA-----PTTPKAPPT 561
Db 1683 FKESAKQMLDPANYGTGMERNWPTPKAEQSLLEDLAGFKELFQTPDHTBESITDDKTKI 1742
QY 562 APKEPAP-----TTPKET--APTTP-----KLTPTPT----- 586
Db 1743 ACKSPPPESMDTPTSTRRPKTPGLGRDIVEELSALKQLTQTHDQVPGDEKGINVFR 1802
QY 587 ---PEKLAP---TTPKEPAPTTP-----EELA-----PTTPEEPTPTTP 619
Db 1803 ERAKQKLDPAASVTGSKRQPRIPKGAQPLEDLAGKELFQTPVCTDKPTTHKTKIAC 1862
QY 620 EEPAP-----TTPKAAA----- 631
Db 1863 RSPQDPVGTPTTFKPSKSLRKADVEEELSALRKRTPSVGKAMDTPKPAVGDEKDKMA 1922
QY 632 -----PTTPKEPA-----PTTPKAPPT 650
Db 1923 FMGTPVOKLDLPNLPFGSKRWPTPKKEKAQALEDLAGFKELFQTPGTDKPTDEKTKIA 1982
QY 651 PKEPAP-----TTPKET-----APTTPK----- 668
Db 1983 CKSPQDPVDTPASTKQPKENLRKADVEEFALFRKRTPSAGKAMDTPKPAVSGEKNIN 2042
QY 669 -----GTAPTTPKEPAPTTPKPPAP-----KELAPTTPKPTSTTSDK- 706

Db 2043 TFFVETPVQKLDLLNLPFGSKRQ--PQTPKEKAQALEDLVGFKELFQTPGHTTEBMTDDKI 2100
QY 707 -----PAP-----TTPKGTAPT----- 718
Db 2101 TEVSCSKPQSPESFKTSRSSKQRLKIPLVKVDKKEEPLAVSKLRTSGTQTHTEPTGDS 2160
QY 719 -----TPK-----TPKEPA----- 732
Db 2161 KSIKAFKESPKQIILDPAAVSGSRRLQTRKAKARALEDLVDPKELFSAPGHTTEBMTID 2220
QY 733 -----PTTPKGTAPTTPKEPAPTTPKPKAPKELAP-----TTTKGTSTTSKPA- 777
Db 2221 KNTKIPCKSPPELTDATSTKRCP-KTRPRKEVKEELSVERLTQTSGETTHHKEPAS 2279
QY 778 -----PTTPKE-----TA 785
Db 2280 GDEGKVLKQRAKKPNPVEEPPSRPRAPKPKAQPLEDLAGFTELSETHQESLTA 2339
QY 786 PTTTPKEPAPTTPKAPPTTPETPTTPPTSEVSTPTTPKPTTIHKSPDESTEPELSAETTP- 844
Db 2340 GKATKIPCESPPLEWDTTASTKRHLRTRVQKVQKEEPSAV-KFTQTSGETTHDADKEPA 2398
QY 845 -----KALENSPKPEGPVIT-----KTPAA--TKPEMTT 871
Db 2399 GEDKGIKALKESAKQTPAPAAVSGSRRRPRAPRESAQALIEDLAGFKDPAGHTEESMT- 2457
QY 872 TAKDITTERDLRTTPE---TTTAPKMTKETATTTEXTITE---SKITATTQVSTTTTQ 924
Db 2458 --DDKTKIPCKSSPELEDATSSKRRPRTRAQKVEVKEELLAVGKLTQTSGETTHDKE 2515
QY 925 -----DTTPFKITLTKTTLAPKVTITTKITITTEIMNKEETAKPKDRA----- 969
Db 2516 PVGEGKGTAFK-----QPAKRNVDADVIGSRROPRAPKPKAQPLEDLASF 2562
QY 970 -----TNSKATTPKQPKTK-----APK-KPT----- 990
Db 2563 QELSQTGHTTEELANGAADSETSAPKQTPDSGKELKISRRVLRAPKPEVPGDVVSTRDPV 2622
QY 991 -----STYKPKTMRVRKPK 1005
Db 2623 KSQSKSNTSLPPLPKRGGGKGSVTGKRLRCMPAPEEIVEELPASKKORVAPRAGKS 2682
QY 1006 TTPT---PRKMTSMPELNPISRIAEAMLOTTTRPNO---TNSKLVEVNFKSEDAAGGA 1058
Db 2683 SEPVVIMKRSURTSAKRIEPAEELNSNDKINKKEHLQDSVPENKGISLRSDQKTEA 2742
QY 1059 EGETHMML-----RPHVFMPEVTPDMDYLPVRVNOGIIINP 1095
Db 2743 BQQITEVFVLAERIEINRNEKKP---MKTSPEND-----IQNP 2777

RESULT 15

A48666
cell proliferation antigen Ki-67, long form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
C:Accession: A48666
R:Schlueder, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerdt
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq-
ins.
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: A48666
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3256 <SCH>
A:Cross-references: UNIPROT.P46013; EMBL.X65550; NID:9415818; PIDN:CAA46519.1; PID:94151
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
P:29-91/Domain: kinase interaction domain homology <KIH>
Query Match 14.5%; Score 855.1; DB 2; Length 3256;
Best Local Similarity 15.9%; Pred. No. 2.5e-09;

Matches 418; Conservative 132; Mismatches 450; Indels 1629; Gaps 100;

```

QY 4 KTLPIYLLLLSVFVIQVSSQDLSSCAGCGEYSRDATC----- 44
Db 600 KTA-----ASSKSQTEVPKRGE--RVATCLQKRVSIRSQHDILOMIC 643
QY 45 -----N 45
Db 644 SKRRSGASEANLIVAKSWADVVVLGAKQTQKVIKHGQPSNMNKRQRRPATPKPVGEVH 703
QY 46 CDYNCHYMECCP-----DFKR----- 62
Db 704 SQFSTGHANSPCTIIIGKAHTEKHVHPARPYVLNFIISNQKMDFKEDLSGIAEMFKTPV 763
QY 63 -----VCTAELS-----CKGRCFESFERGECDDAOCKYKCCPDYESFCAEV--- 107
Db 764 KEQPOLSTSTCHIAISENELLGQFGTDSGEE-----PLLPSTSEFGNGVFFS 812
QY 108 -HNPTSPSSK-KAPPP-----SGASQIKST 132
Db 813 AQNAKQPSDKCSASPPPLRQRCIRENGVAKTPRNTYKMTSLETETEPSTV-ST 871
QY 133 TKRS-----PKPPNKKKT----- 145
Db 872 VNRSGRSTFRNIQKLPVESKSEETNTEIVECILKRGQKATILQORREGEKMEIERPPET 931
QY 146 -KKVIESEITEKVDXNKNRT-KKPTPKPPVDEAGSLDNGDFKVTTPDT----- 196
Db 932 YKENIELKENDXKAMKASRTWGOKCAPMSDLTD-----LKSLPDTELMKDTA 980
QY 197 -----STTQHNKSVSPKITTAKINPRPSLPP-----NSDTS-----KETS 233
Db 981 RQNLILQTDHAKAPKSEKGIK--MPCQSLOPEPINTPTHTKQOLKASLGKGVKEEL 1038
QY 234 LVNKEITVETKETTNTKQSTDSKE-----KTSAKETO 269
Db 1039 LAVGFTRT-SGETHTHREBPAGDGKSIRTFKESPKQLDPAARVTGMKKWPRTPKEBAQ 1097
QY 270 SIEKTSAKDLAPTSKVLAKPTPKAE-----TTTK----- 298
Db 1098 SLE-----DLAGFKELFQTPGSEBSMTDEKTKIACKSPPPESVDTPSTKQWPKRSLR 1152
QY 299 -----GPALTTPKEP-----TPT----- 311
Db 1153 KADVEEFELALRKLTPSAGKAMLTTPKAGGDEKOIKAFMGTPVQKLDLAGLPGSKRQLQ 1212
QY 312 TPKEPAST-----TPKEP-----TPTTIKSAPT-- 334
Db 1213 TPKEAQALEDLAGFKELFQTPGHTHEELVAAGKTTKIFCDSPQSDPVDTPSTKQPKRS 1272
QY 335 -----TPKEPAP-----TTT 344
Db 1273 IRKADVEGELLACRLNMPKAGKAMHTPK--PSVGEEDKIIIFVGTPVQKLDLTENLTGS 1329
QY 345 KSAPTTPKEPAPT-----TTKEPAPTTPKEPAPTTPKEPA 379
Db 1330 KRRPOTPKEEAQALEDLTGFKELFQTPGHTEEVAAGKTTKMPCESSPPESADT----- 1383
QY 380 PTTTKSAPTTPKE----- 392
Db 1384 PTSTRRQKPTPLEKRDVQKELSALKKLTQTSGETHTTKVPGEDKSNAPRETAQKLD 1443
QY 393 -PAPTTPKPAAPTTPKEA-----PTTPK-----EPTPT- 420
Db 1444 PAASVTGSKRHPKT-KEAQAQLEDLAGWKELFQTPVCTDKXTTHEKTTKIACRSQDPVD 1502
QY 421 TPKEPAPTTPKE-----PAPT----- 435
Db 1503 TPTSSKPSKSLRKVDVEEFELALRKLTPSAGKAMHTPKPAVSGEKNIAFYAFMTPGVQKL 1562
QY 436 -----TPKEPA-----PTA 444
Db 1563 DLTENLTGSKRLQTPKEAQAQLEDLAGFKELFQTPRGHTESMTNDKTAACKVACKSQPDL 1622

```

```

QY 445 PKKPA-----PTTPKEPAPT----- 459
Db 1623 DKNPASKRRLKTSLGKGVKEELLAVGKLTQTSGETHTHTHEPTGDGKSMKAFMESPKQ 1682
QY 460 -----TPKEPA-----PTTKEPSPTTPK-----E 479
Db 1683 ILDSAASLGSKQKLTTPGKSEVPEDLAGFIELFQTPSHTKS-SMTNEKTKVSVRASQ 1741
QY 480 P-----APTTPKSAP----- 489
Db 1742 PDLVDTPSSKPOPKRSLKADTEEEFLAPRKQTPSAGKAMHTPKPAVGEEDKINTFLGT 1801
QY 490 -----TTTKEPA-----PT-----TTK-----SA 503
Db 1802 PVOKLDOPGNLPGSNRRRLQTRKEKAQALBELTGFRELFPQTPCTDNPDADEKTTKILCKS 1861
QY 504 PTTPEPSPTTPKE-----PA----- 519
Db 1862 PQSDPADTPNTKQPKRSILKADVEEFLAPRKLTPSAGKAMHTPKAAVGEEDKINTFV 1921
QY 520 -----PTTPKEPAPT-----TP----- 531
Db 1922 GTPVEKLDLGNLPGSKRRFPQTPKEKAKALEDLAGFKELFQTPGHTESMTDDKITEVSC 1981
QY 532 KKPAPTTPKEPA----- 543
Db 1982 KSPQDPVKPTPTSSKQRLKISLGKGVKEEVLVPGKLTQTSKGTQTHRETAGDGKSIKA 2041
QY 544 -----PTTPKEPA-----PTTTKPAAPT 561
Db 2042 FKESAKQMLDPANYGTGMRWPRTPKEEAQSLLEDLAGFKELFQTPDHTBESTDDKTKI 2101
QY 562 APKEPAP--TPPKET--APTTP-----KLTPTT----- 586
Db 2102 ACKSPPEMDPTSTRRPKTPLGKRDIVBELSALKQLTQTHTDKVPCEDEKGINVFR 2161
QY 587 ---PEKLAP---TTPEKPAPTTP-----EELA-----PTTPEPTPTTP 619
Db 2162 ETAKQLDPAASVTGSKRQRPTRPKGAQPLEDLAGLKELFQTPVCTDKPTTHEKTTKIA 2221
QY 620 EEPAP-----TTPKAAA----- 631
Db 2222 RSPQDPVGTPTIFKPSKRSRKADVBEESIALRKRTPSVGKAMDTPKPAGDEKDKMA 2281
QY 632 -----PNTPKPEA-----PTTPKEPAPT 650
Db 2282 FMGTPVQKLDLPGNLPGSKRWQTPPKKAQALEDLAGFKELFQTPGTDKPTTDEKTKIA 2341
QY 651 PKPEPAP--TTPKET-----APTTPK----- 668
Db 2342 CKSPQDPVDTPASTQRPKRNLRKADVBEELALRKLTPSAGKAMDTPKPAVSBKNTN 2401
QY 669 -----GTAPTTLKSPAPTTPKKPAP-----KELAPTTPKPTSTTSK- 706
Db 2402 TFWETPVQKLDLGNLPGSKRQ--POTPKKABALEDLVGFELFQTPGHTESMTDDKI 2459
QY 707 -----PAP-----PTPKGTAPT----- 718
Db 2460 TEVSCKSPQESFKTSRSSKQRLKIPLVKVDKKEPFLAVSKLRTSGETTQTHTEPTGDS 2519
QY 719 -----TPK--BPAPT-----TPKEPA----- 732
Db 2520 KSIKAPKESFKQLDPAASVTGSRQRLRTRKKAQALEDLVDFKELFSPAGHTESMTID 2579
QY 733 -----PTPKGTAPTTLKSPAPTTPKKPAPKELAP-----TTTKGPTSTTSKPA- 777
Db 2580 KNTKIPCKSPPPBELTDTATSTKRCP-KTRPRKEVKEELSAVERLTQTSQSTHHEKPA 2638
QY 778 -----PTTPKE-----PTA 785
Db 2639 GDEGKVLQRAKKKNPVVEEPPSRRRPRAPKKAQPLEDLAGFTELSETSGHTQESLTA 2698

```

```
QY 786 PTPKBPATTPKBPATTPPTTSEVSTPTTKEPTTIHKSPDESTPESAEPTP- 844
Db 2699 GKATKIPCESPLEVVDTTASTKRHLRTVQVQVKEEPSAV-KFTQTSGETTDADKEPA 2757
QY 845 -----KALENSPKKEGVPTT-----KTPAA--TKPEMTT 871
Db 2758 GEDKGKALKESAKQTPAASVTSRRPRAPRESAQAIEDLAGFKDPAAGHTEESNT- 2816
QY 872 TAKDKTTERDLRTPE---TTTAAPKQWKETATTEKTE---SKITATTQVTSITQ 924
Db 2817 --DDKTKIPCKSSPELEDATSSKRRPRTRAQKVEVKEELLAVGKLQTSGETTHDK 2874
QY 925 -----DTTPFKITTLKTLAPKVTTKTITTEIMNKPEETAAPKORA----- 969
Db 2875 PVGEGKGTKAFK-----OPAKRNVDAEDVIGSRQPRAPKEKAQPLEDLASF 2921
QY 970 -----TNSKATTPKQKPTK-----APK-KPT----- 990
Db 2922 QELSQTPGHTBELANGAADSFTSAPKQTPDSGKPLKISRRVLRAPKVEPVGDVVSTRDPV 2981
QY 991 -----STKKPKTMPVRKPK 1005
Db 2982 KQSKSNTSLPLPFKRGKGDSVGTGKRLRCMPAPEIIVEELPASKKQKVAPRARGKS 3041
QY 1006 TTPT---PRKMTSTWPELNPTSRIAEAMLQTTTRPNQ---TPNGSKLVEVNPKSEDAGGA 1058
Db 3042 SEPVVIMKESLRTSAKRIEPABELNSNDMKNKEHKLODSVPENKGISLSRRQDKTEA 3101
QY 1059 EGETHMLL-----RPHVFMPEVTPDMVYLPVPVNOGIIINP 1095
Db 3102 EQQITEVFVLABRIEINRNEKFP---MKTSPEMD-----IQNP 3136
```

Search completed: October 13, 2004, 11:58:11
Job time : 68.3652 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 84.5733 Seconds
(without alignments)
4449.477 Million cell updates/sec

Title: SEQ1-D
Perfect score: 576
Sequence: 1 MAWKLPYLLLSVFIQ.....DMDYLPVPGIINPMLS 1049

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 2002273 seqs, 35872929 residues
Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 23Sep04: *
1: Geneseqp1980s: *
2: Geneseqp1980s: *
3: Geneseqp2000s: *
4: Geneseqp2001s: *
5: Geneseqp2002s: *
6: Geneseqp2003as: *
7: Geneseqp2003bs: *
8: Geneseqp2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5546.9	99.5	1299	4 AAM24322	Human EST
2	5546.9	99.5	1404	2 AAR26049	MSF precu
3	5546.9	99.5	1404	4 AAB29773	Human meg
4	5546.9	99.5	1404	4 AAB60568	Human meg
5	5546.9	99.5	1404	8 ADP98014	Human meg
6	5546.9	99.5	1415	4 AAU32262	Novel hum
7	5523.9	99.1	1404	7 ADK65839	Angiogene
8	5384.3	96.6	1320	7 ADK65819	Angiogene
9	5358.7	96.1	1270	8 ADK67911	Human ext
10	5344.6	95.9	1311	8 ADK67912	Human ext
11	2920	52.4	546	4 ABUS3252	Human tes
12	2888	51.8	551	4 ABUS3253	Human tes
13	2880.1	51.7	902	4 AAB29778	Human Mef
14	2757	49.4	513	4 ABUS3254	Human tes
15	1730.1	31.0	452	4 AAR80041	Human meg
16	1409.9	25.3	538	5 AAU18834	3' cartil
17	1346.8	24.2	5179	6 AAM24516	C899P pre
18	1346.8	24.2	5179	6 ABP55365	Human col
19	1346.8	24.2	5179	6 ABO07258	Human p53
20	1346.8	24.2	5179	7 ADD48091	Human Pro
21	1346.8	24.2	5179	7 ADP44998	Human HAP
22	1174.4	21.1	232	5 AAU11261	Human HAP
23	1128.7	20.2	1664	2 AAW43106	C. thermo
24	1099.6	19.7	717	4 ABUS3144	Human tes
25	1097	19.7	214	4 ABUS3255	Human tes

26	1088.5	19.5	8991	6 ABU08487	S. pneumo
27	1068	19.2	1795	4 ABB69806	Drosophil
28	1042.8	18.7	763	3 AAG38942	Arabidops
29	1039.7	18.6	1049	4 ABB61364	Arabidops
30	1020.5	18.3	2284	4 ABB71434	Drosophil
31	987.9	17.7	770	4 ABUS3141	Human tes
32	981	17.6	778	4 ABUS3143	Human tes
33	978.7	17.6	5703	8 ADL23265	Human MUC
34	975.9	17.5	2112	4 ABB60403	Drosophil
35	974.4	17.5	692	4 ABUS3155	Human tes
36	964.3	17.3	4315	5 ABP43908	MUC5B par
37	964.1	17.3	717	4 ABUS3150	Human tes
38	964.1	17.3	717	4 ABUS3149	Human tes
39	964.1	17.3	717	4 ABUS3151	Human tes
40	964.1	17.3	717	4 ABUS3145	Human tes
41	964.1	17.3	717	4 ABUS3148	Human tes
42	964.1	17.3	717	4 ABUS3147	Human tes
43	964.1	17.3	717	4 ABUS3146	Human tes
44	952.1	17.1	745	4 ABUS3142	Human tes
45	951.6	17.1	1538	7 ADI21202	Novel hum

ALIGNMENTS

RESULT 1
AAM24322
ID AAM24322 standard; protein; 1299 AA.

AC AAM24322;
DT 12-OCT-2001 (first entry)
DE Human EST encoded protein SEQ ID NO: 1847.

KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;
KW gene therapy; nutrition.

XX Homo sapiens.

XX WO200154477-A2.

XX PD 02-AUG-2001.

XX PF 25-JAN-2001; 2001WO-US002687.

XX PR 25-JAN-2000; 2000US-00491404.

XX PR 17-JUL-2000; 2000US-00617746.

XX PR 03-AUG-2000; 2000US-00631451.

XX PR 15-SEP-2000; 2000US-00663870.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

XX PI Cao Y, Drmanac RA, Zhang J, Werhman T;

XX DR WPI; 2001-476164/51.

XX DR N-PSDB; AAR98981.

XX PT Isolated polypeptide for treatment of diseases, diagnostics, raising

XX PT antibodies and research use.

XX PS Claim 20; Page 1198-1201; 12/5pp; English.

XX CC The present invention provides the protein and coding sequences of novel proteins from a variety of organisms, including human, dog, cat, horse, cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea urchin and tomato. These were derived from expressed sequence tags (ESTs) from the organism of interest. They can be used in diagnostics, forensics, gene mapping, identification of mutations, to assess biodiversity and for nutritional purposes. The present sequence is a

```

CC protein of the invention
XX
SQ Sequence 1299 AA;

Query Match          99.5%; Score 5546.9; DB 4; Length 1299;
Best Local Similarity 92.0%; Pred. No. 8.8e-152;
Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

QY 1 MAWKTLPIYLLLLLSVFVIQQVSSQ----- 25
DB 1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSGACGCGEYSRDATCNCYDNCQHYMECCPDF 60
QY 26 -----ELCKGRCRFPESFGRGECDCDAQCKYDKCCPDYDFCA----- 64
DB 61 KRVCTAELSKGRCRFPESFGRGECDCDAQCKYDKCCPDYDFCAEYVHNPTSPSSKKAP 120
QY 65 -----BEHSVSENBQSSSSSSSSSSSSSSSIW 89
DB 121 PPSGASQTIKSTKSPKPPNKKTKVIESBEITEHSVSENBQSSSSSSSSSSSSSIW 180
QY 90 KIKSKNSAANRELOKLVKDNKNKNTKKKPTPKPPVVDGAGSLDNGDFKVTTPDTST 149
DB 181 KIKSKNSAANRELOKLVKDNKNKNTKKKPTPKPPVVDGAGSLDNGDFKVTTPDTST 240
QY 150 TQHNKVSTSPKLTAKPINRPSLPNSDTSKETSLSLVNKTETVETKETTNTKQTSIDG 209
DB 241 TQHNKVSTSPKLTAKPINRPSLPNSDTSKETSLSLVNKTETVETKETTNTKQTSIDG 300
QY 210 KEKTTSAKETQSIKTSANDLAPTSLKVLAKPTPKAETTTKGPALTTPKPTPTPKBPAS 269
DB 301 KEKTTSAKETQSIKTSANDLAPTSLKVLAKPTPKAETTTKGPALTTPKPTPTPKBPAS 360
QY 270 TTPKPTPTTIISAPTTKEPAPTTTKSAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEP 329
DB 361 TTPKPTPTTIISAPTTKEPAPTTTKSAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEP 420
QY 330 APTTTKSAPTTKEPAPTTPKKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTPK 389
DB 421 APTTTKSAPTTKEPAPTTPKKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTPK 480
QY 390 EPAPTAPKAPAPTTKEPAPTTPKAPPTTKKPSPTTPKEPAPTTTKSAPTTKEPAPT 449
DB 481 EPAPTAPKAPAPTTKEPAPTTPKAPPTTKKPSPTTPKEPAPTTTKSAPTTKEPAPT 540
QY 450 TTKSAPTTPKPSPTTPKEPAPTTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAP 509
DB 541 TTKSAPTTPKPSPTTPKEPAPTTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAP 600
QY 510 APTAPKEPAPTTPKETAPTTPKLTPTTPEKLAPTTPEKAPPTTPEELAPTTPEEPTPTT 569
DB 601 APTAPKEPAPTTPKETAPTTPKLTPTTPEKLAPTTPEKAPPTTPEELAPTTPEEPTPTT 660
QY 570 PEEPAPTTPKAAPNTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPT 629
DB 661 PEEPAPTTPKAAPNTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPT 720
QY 630 APTTPKPKAPKELAPTTTKETSTTSKAPPTPKGTAPTTPKEPAPTTTPKEPAPTTTPKG 689
DB 721 APTTPKPKAPKELAPTTTKETSTTSKAPPTPKGTAPTTPKEPAPTTTPKEPAPTTTPKG 780
QY 690 TAPPTLKAPAPTTPKKAPKELAPTTTKGPTSTTSKAPPTPKGTAPTTPKEPAPTTTPKEPAPT 749
DB 781 TAPPTLKAPAPTTPKKAPKELAPTTTKGPTSTTSKAPPTPKGTAPTTPKEPAPTTTPKEPAPT 840
QY 750 KPAPTTPTPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALENSPKPEGVPT 809
DB 841 KPAPTTPTPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALENSPKPEGVPT 900
QY 810 TKTPAATPKEMTTAKDKITRDLTTPETTTAAPKMTKETATTTKTTESKITATTQV 869
DB 901 TKTPAATPKEMTTAKDKITRDLTTPETTTAAPKMTKETATTTKTTESKITATTQV 960
QY 870 TSTTTQDTPPKITTLTKTLTAPKVTTTKTITTEIMNKPBEETAKPKDRATNSKATTPK 929

```

```

DB 961 TSTTTQDTPPKITTLTKTLTAPKVTTTKTITTEIMNKPBEETAKPKDRATNSKATTPK 1020
QY 930 PQKPTKAPKPTSTKKPKTMPVRVKETPTPKMTSTMPELNPTSRIAEAMLOTTTREN 989
DB 1021 PQKPTKAPKPTSTKKPKTMPVRVKETPTPKMTSTMPELNPTSRIAEAMLOTTTREN 1080
QY 990 QTNSKLVEYNPKSDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPVRVNOGIIINPMLS 1049
DB 1081 QTNSKLVEYNPKSDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPVRVNOGIIINPMLS 1140

RESULT 2
AAR26049
ID AAR26049 standard; protein; 1404 AA.
XX
AC AAR26049;
XX
DT 25-MAR-2003 (revised)
DT 02-FEB-1993 (first entry)
XX
DE MSF precursor.
XX
KW Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
KW stability; proteolytic cleavage; adhesion; alternative splicing.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Region 1..26 /label= Exon_I
FT Region 26..67 /label= Exon_II
FT Region 67..107 /label= Exon_III
FT Region 107..157 /label= Exon_IV
FT Region 157..200 /label= Exon_V
FT Region 200..1141 /label= Exon_VI
FT Region 1166..1212 /label= Exon_VIII
FT Region 1213..1266 /label= Exon_IX
FT Region 1266..1331 /label= Exon_X
FT Region 1331..1373 /label= Exon_XI
FT Region 1373..1404 /label= Exon_XII
FT Region 1411..1166 /label= Exon_VII
XX
PN WO9213075-A1.
XX
PD 06-AUG-1992.
XX
PF 17-JAN-1992; 92WO-US0000433.
XX
PR 18-JAN-1991; 91US-00643502.
PR 10-SEP-1991; 91US-00757022.
XX
PA (GEMY ) GENETICS INST INC.
XX
PI Turner K, Clark SC, Jacobs K, Hewick RM, Gesner TG;
PI WPI; 1992-284660/34.
XX
DR N-PSDB; AAQ27223.
XX
PT New human mega-karyocyte stimulating factors - for treating immune
PT deficiencies, cancer, exposure to radiation or drugs, bacterial and viral
PT infections, etc.

```


CC	The invention relates to a method of treating osteoarthritis via the
CC	administration of a composition comprising the camptodactyly-arthropathy-
CC	coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
CC	The composition may further comprise a local anaesthetic. The composition
CC	of the invention may be administered via intra-articular or intravenous
CC	injection. The human CACP prorein is identified in the invention as being
CC	megakaryocyte stimulating factor (MSF). The gene encoding CACP protein
CC	(MSF) is located on chromosome 1q25-31, and mutations in this gene are
CC	responsible for the heritable disorder camptodactyly- arthropathy-coxa
CC	vara-pericarditis, in which patients have synovial hyperplasia without
CC	evidence of inflammation. CACP protein (MSF) acts as a synovium
CC	lubricant, and can be used to lubricate tissue and joints in the
CC	treatment of osteoarthritis. The composition may be applied to reduce the
CC	symptoms of osteoarthritis (e.g., joint pain, loss of range of movement
CC	or joint damage). The present sequence represents human megakaryocyte
CC	stimulating factor (MSF, CACP protein). Note: This sequence is not given
CC	in its entirety in figure 4 of the specification, although a GenBank
CC	accession number was given. This sequence was therefore obtained from
CC	GenBank (U70316)
XX	
SQ	Sequence 1404 AA;
	Query Match 99.5%; Score 5546.9; DB 4; Length 1404;
	Best Local Similarity 92.0%; Pred. No. 9.7e-152;
	Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2
QY	-
DG	1 MAWKTLPIYLILLLSVFVIOOVSSQ----- 25
DG	1 MAWKTLPIYLILLLSVFVIOOVSSQDLSACAGCGEGYSRDATCNCDYNQCQHYMECCPDF 60
QY	26 -----ELSCGRCPESFERGREDCDAQCKKYDKCCPDYESPCA----- 64
DG	61 KRVCTAELSCGRCPESFERGREDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSKKAP 120
QY	65 -----REHSVSENQCESSSSSSSSSSSSSSTIIV 89
DG	121 PFGASQTIKSTTKGSPPNKXKTKKVIESEITEEHSVSENQCESSSSSSSSSSSSTIIV 180
QY	90 KTSSKNGANRELQKLKVONKNRNTKKTPKPWDEAGSGLDNGDFKVTTPDTST 149
DG	181 KTSSKNGANRELQKLKVONKNRNTKKTPKPWDEAGSGLDNGDFKVTTPDTST 240
QY	150 TQHNKVSTSPKITTAKPINRPSPSPNSDTSKETSLTVNKETTVEKTTNNKQISTGD 209
DG	241 TQHNKVSTSPKITTAKPINRPSPSPNSDTSKETSLTVNKETTVEKTTNNKQISTGD 300
QY	210 KEKTTSAKETQSIEKTSADLAPTSKVLAKPTPKAETTTKGPAITTPKEPTPTTPKEPAS 269
DG	301 KEKTTSAKETQSIEKTSADLAPTSKVLAKPTPKAETTTKGPAITTPKEPTPTTPKEPAS 360
QY	270 TTPKEPTPTTIKSAPTTTPEPAPTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 329
DG	361 TTPKEPTPTTIKSAPTTTPEPAPTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 420
QY	330 APPTTKSAPTTTPEPAPTTTPKKAPPTTPEPAPTTTKEPPTTTPKEPAPTTTKEPAPTTTK 389
DG	421 APPTTKSAPTTTPEPAPTTTPKKAPPTTPEPAPTTTKEPPTTTPKEPAPTTTKEPAPTTTK 480
QY	390 EPAPTAPKKAPPTTPEPAPTTTPEPAPTTTTPKEPSPTTPKEPAPTTTTSAPTTTKEPAPT 449
DG	481 EPAPTAPKKAPPTTPEPAPTTTPEPAPTTTTPKEPSPTTPKEPAPTTTTSAPTTTKEPAPT 540
QY	450 TTTSAPTTTPEPSPTTTPKEPAPTTTPEPAPTTTPKKAPPTTPEPAPTTTKEPAPTTTKKP 509
DG	541 TTTSAPTTTPEPSPTTTPKEPAPTTTPEPAPTTTPKKAPPTTPEPAPTTTKEPAPTTTKKP 600
QY	510 APTAPKEPAPTTPKETAPTTPKLTPTTPEKLAPTTTPKEKAPPTTPEBLAPTTTPEEBPTPT 569
DG	601 APTAPKEPAPTTPKETAPTTPKLTPTTPEKLAPTTTPKEKAPPTTPEBLAPTTTPEEBPTPT 660
QY	570 PEPEAPTTTCAAAPNTPKEPAPTTTPEPAPTTTPEKAPPTTTPKETAPTTTGGTAPTTLKGP 629
DG	661 PEPEAPTTTCAAAPNTPKEPAPTTTPEPAPTTTPEKAPPTTTPKETAPTTTGGTAPTTLKGP 720


```

QY 301 TPKEPAPTTTKBPAPTTTPKEPAPTTTKBPAPTTTKSAPTTTKBPAPTTTKBPAPTTTPKEP 360
Db 258 TPKEPAPTTTKBPAPTTTPKEPAPTTTKBPAPTTTKSAPTTTKBPAPTTTKBPAPTTTPKEP 317
QY 361 APPTPKPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTAPKBPAPTTTPKEPAPTTTPKEPAPTTT 420
Db 318 APPTPKPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTAPKBPAPTTTPKEPAPTTTPKEPAPTTT 377
QY 421 KPSPPTTPKEPAPTTTKSAPTTTKBPAPTTTKSAPTTTKBPAPTTTPKPSPTTPKEPAPTTTPKEPAPT 480
Db 378 KPSPPTTPKEPAPTTTKSAPTTTKBPAPTTTKSAPTTTKBPAPTTTPKPSPTTPKEPAPTTTPKEPAPT 437
QY 481 TPKKPAPTTTPKEPAPTTTPKEPAPTTTKBPAPTTAPKBPAPTTTPKEPAPTTTPKPLTPTTEK 540
Db 438 TPKKPAPTTTPKEPAPTTTPKEPAPTTTKBPAPTTAPKBPAPTTTPKEPAPTTTPKPLTPTTEK 497
QY 541 LAPTTPEKAPTTPELAPTTPEEPTTPPEBPAPTTPKAAAPNTPKEPAPTTTPKEPAPT 600
Db 498 LAPTTPEKAPTTPELAPTTPEEPTTPPEBPAPTTPKAAAPNTPKEPAPTTTPKEPAPT 557
QY 601 TPKEPAPTTPKETAPTTPKGTAPTTLKBPAPTTPKKBPAPKELAPTTTKEPTSTSDKPAP 660
Db 558 TPKEPAPTTPKETAPTTPKGTAPTTLKBPAPTTPKKBPAPKELAPTTTKEPTSTSDKPAP 617
QY 661 TTPKGTAPTTTPKEPAPTTTPKGTAPTTLLKEPAPTTTPKKPAPKELAPTTTKGPT 720
Db 618 TTPKGTAPTTTPKEPAPTTTPKGTAPTTLLKEPAPTTTPKKPAPKELAPTTTKGPT 677
QY 721 STSDKPAPTTPKETAPTTTPKEPAPTTPKKBPAPTTPETPPTTSVSTPTTKETPTIHK 780
Db 678 STSDKPAPTTPKETAPTTTPKEPAPTTPKKBPAPTTPETPPTTSVSTPTTKETPTIHK 737
QY 781 SPDESTPELSAPTPKALENSPKPGVPTTKTPAATKPEMTTAKDKTTERDLATTPETT 840
Db 738 SPDESTPELSAPTPKALENSPKPGVPTTKTPAATKPEMTTAKDKTTERDLATTPETT 797
QY 841 TAAPKPKETATTTKTTESKITATTQVSTSTTQDTTPFKITLTKTTTLAPKVTTTKT 900
Db 798 TAAPKPKETATTTKTTESKITATTQVSTSTTQDTTPFKITLTKTTTLAPKVTTTKT 857
QY 901 ITTITIMNKPETAKPKORATNSKATTPKPKPTKAPKPTSTKPKTMVVRPKTTPT 960
Db 858 ITTITIMNKPETAKPKORATNSKATTPKPKPTKAPKPTSTKPKTMVVRPKTTPT 917
QY 961 PKMTSTMPELNPTSRIAEAMLQTTTRNQTPNSKLVBNPKSBDAGGAGETPHMLLRP 1020
Db 918 PKMTSTMPELNPTSRIAEAMLQTTTRNQTPNSKLVBNPKSBDAGGAGETPHMLLRP 977
QY 1021 HVFMEVTPDMDYLPRVNPQGIINPMLS 1049
Db 978 HVFMEVTPDMDYLPRVNPQGIINPMLS 1006

RESULT 10
ID ADK67912
XX ADK67912 standard; protein; 1311 AA.
AC ADK67912;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human extracellular messenger (EXMES) polypeptide.
XX
KW Human; extracellular messenger; EXMES; respiratory-gen.; antiallergic;
KW antitachymatic; antiinflammatory; antidiabetic; neuroprotective;
KW muscular-gen.; antiarthritic; osteopathic; hepatotropic; antiparasitic;
KW virucide; fungicide; antiparasitic; protozoacide; antihelminthic;
KW cytostatic; gene therapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers

```

```

FT Peptide
FT 1..30
FT /label= Signal_peptide
FT /note= "Spans residues 1 to 18, 20, 21, 24, 29 or 30
XX according to identification method"
XX
XX W02004013292-A2.
XX
XX 12-FEB-2004.
XX
XX 30-JUL-2003; 2003W0-US024084.
XX
XX 02-AUG-2002; 2002US-0400810P.
XX
XX 19-SEP-2002; 2002US-0412197P.
XX
XX 04-OCT-2002; 2002US-0416004P.
XX
XX 08-NOV-2002; 2002US-0424862P.
XX
XX (INCY-) INCYTE CORP.
XX
XX Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP;
XX Richardson TW, Emerling BM, Lindquist BA, Chawla NK, Ramkumar J;
XX Lee SY;
XX
XX WPI; 2004-157116/15.
XX
XX N-PSDB; ADK67917.
XX
XX New extracellular messengers and nucleic acids, useful for diagnosing,
XX treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes
XX mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or
XX autoimmune thyroiditis.
XX
XX Claim 60; SEQ ID NO 5; 165pp; English.
XX
XX The present sequence is that of novel human extracellular messenger
XX (EXMES) Incyte ID NO: 7513018CD1 polypeptide. The protein shows homology
XX to human megakaryocyte stimulating factor. The invention provides EXMES
XX polynucleotides and polypeptides, as well as expression vectors, host
XX cells, antibodies, agonists and antagonists, and methods for diagnosing,
XX treating or preventing disorders associated with aberrant expression of
XX EXMES, especially autoimmune and inflammatory disorders, cell
XX proliferative disorders and endocrine disorders, e.g. adult respiratory
XX distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's
XX disease, diabetes mellitus, myasthenia gravis, osteoarthritis,
XX osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal,
XX parasitic, protozoal or helminthic infections, cancers, autoimmune
XX thyroiditis, cretinism, Plummer's disease or thyroid carcinoma.
XX Embodiments also provide methods for using the purified EXMES and/or
XX their encoding polynucleotides for facilitating the drug discovery
XX process, including determining of efficacy, dosage, toxicity and
XX pharmacology, and for investigating the pathogenesis of diseases and
XX medical conditions.
XX
XX Sequence 1311 AA;

```

```

Query Match 95.9%; Score 5344.6; DB 8; Length 1311;
Best Local Similarity 92.3%; Pred. No. 6.1e-146;
Matches 1006; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNDCYNQHYMECCPDF 60
QY 26 -----ELSCKGRCFESFERGECDCDAQCKKYDKCCPDYESFCAEHSVSENQSSSS 79
Db 61 KRVCYAEKSKGRCFESFERGECDCDAQCKKYDKCCPDYESFCAE----- 106
QY 80 SSSSSSTTWIKSKNSKNSAANRELQKKLVKNQKNRKKKPTKPPVVDAGSLDNGD 139
Db 107 -----VKDNKNRKKKPTKPPVVDAGSLDNGD 137
QY 140 FKVTPTDSTTQHNKYSTSPKITTAKPINPRSLPNSDTSKETSITVNTKETT 199
Db 138 FKVTPTDSTTQHNKYSTSPKITTAKPINPRSLPNSDTSKETSITVNTKETT 197

```



```
QY 200 TTNKQSTGKKTSAKETQSIKTSAXDLAPTQSVLAKPTPKAETTTKGALATPKPEP 259
Db 198 TTNKQSTGKKTSAKETQSIKTSAXDLAPTQSVLAKPTPKAETTTKGALATPKPEP 257
QY 260 TPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTPKEPAPTTTKKEPAPTTTPK 319
Db 258 TPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTPKEPAPTTTKKEPAPTTTPK 317
QY 320 EPAPTTTKPAPTTTKSAPTTTPKEPAPTTPKAPATTPTPKEPAPTTPKBPTTPTPKEPAPT 379
Db 318 EPAPTTTKPAPTTTKSAPTTTPKEPAPTTPKAPATTPTPKEPAPTTPKBPTTPTPKEPAPT 377
QY 380 TKEPAPTTTPKEPAPTTAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTSA 439
Db 378 TKEPAPTTTPKEPAPTTAPKAPATTPTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTSA 437
QY 440 PTTTKPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 499
Db 438 PTTTKPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 497
QY 500 EPAPTTTKPAPTTAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTPEELAP 559
Db 498 EPAPTTTKPAPTTAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTPEELAP 557
QY 560 TTPPEEPTPTTPKEPAPTTTPKAAAPNTPKAPATTPTPKEPAPTTTPKEPAPTTTPKETAPTTPK 619
Db 558 TTPPEEPTPTTPKEPAPTTTPKAAAPNTPKAPATTPTPKEPAPTTTPKEPAPTTTPKETAPTTPK 617
QY 620 GTAPTTLKEPAPTTPKKAPKELAPTTTKETPTSTSDKAPATTPKGTAPTTPKEPAPTTTPK 679
Db 618 GTAPTTLKEPAPTTPKKAPKELAPTTTKETPTSTSDKAPATTPKGTAPTTPKEPAPTTTPK 677
QY 680 KEPAPTTTPKGTAPTTLKEPAPTTTPKAPKELAPTTTKGTPTSTSDKAPATTPKGTAPTTP 739
Db 678 KEPAPTTTPKGTAPTTLKEPAPTTTPKAPKELAPTTTKGTPTSTSDKAPATTPKGTAPTTP 737
QY 740 PKPAPTTTPKAPATTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 799
Db 738 PKPAPTTTPKAPATTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 797
QY 800 NSPKPVGPTTKTAPKAPKEMTTAKDKTTERDLRTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 859
Db 798 NSPKPVGPTTKTAPKAPKEMTTAKDKTTERDLRTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 857
QY 860 SKITATTTQVSTTTQDTTPFKITLTKTTTLAPKVTITTKTITTTTEIMNKPEETAKPKDR 919
Db 858 SKITATTTQVSTTTQDTTPFKITLTKTTTLAPKVTITTKTITTTTEIMNKPEETAKPKDR 917
QY 920 ATNSKATTPKQKPTKAPKPTSTKPKTMPVRKPKTTTPTRKMTSTMPNLNPTSRITAE 979
Db 918 ATNSKATTPKQKPTKAPKPTSTKPKTMPVRKPKTTTPTRKMTSTMPNLNPTSRITAE 977
QY 980 AMLQTTTRNQTPNSKLVNPKSEDAGGAGETPHMLLRPHVFWPEVTPDMDYLPRVFN 1039
Db 978 AMLQTTTRNQTPNSKLVNPKSEDAGGAGETPHMLLRPHVFWPEVTPDMDYLPRVFN 1037
QY 1040 QGIINPMLS 1049
Db 1038 QGIINPMLS 1047
```

RESULT 11

ABU53252

ID ABU53252 standard; protein; 546 AA.

XX

AC ABU53252;

XX

DT 14-APR-2003 (first entry)

XX

DE Human testes-derived DXF2phtes3_4ol9 homologue #1.

XX

KW Human; gene therapy; vaccine; disease treatment; detection.

XX

OS Homo sapiens.
XX WO200112659-A2.
PN
XX
PD 22-FEB-2001.
XX
PF 18-AUG-2000; 2000WO-IB001496.
XX
PR 18-AUG-1999; 99US-0149499P.
PR 28-SEP-1999; 99US-0156503P.
XX
PA (GBHU-) GERMAN HUMAN GENOME PROJECT.
XX
XX Wiemann S;
XX WPI; 2001-327840/34.
DR
XX
PT Nucleic acids having the sequences of clones isolated from libraries of
PT different human tissues, useful in recombinant DNA methodologies.
XX
XX Example III; Page 892; 1095pp; English.
XX
CC This invention describes novel polynucleotides and polypeptides isolated
CC from human cDNA libraries which can be used for gene therapy or in
CC vaccines. The polynucleotides of the invention and antibodies encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. The products of the
CC invention may also be used to identify modulators of expression and
CC activity and to down regulate expression and activity. The antibodies of
CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence represents a homologue
CC of a polypeptide described in the disclosure of the invention
XX
SQ Sequence 546 AA;

Query March 52.4%; Score 2920; DB 4; Length 546;

Best Local Similarity 100.0%; Pred. NO. 1.5e-16;

Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 397 KKPAPTTPKBPAPTTTPKEPAPTTTKESPTTPKEPAPTTTKSAPTTTKPAPTTTKSAPT 456
Db 1 KKPAPTTPKBPAPTTTPKEPAPTTTKESPTTPKEPAPTTTKSAPTTTKPAPTTTKSAPT 60
QY 457 TPKEPSTTTTPKEPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTKKAPAPKE 516
Db 61 TPKEPSTTTTPKEPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTKKAPAPKE 120
QY 517 PAPTPKETAPTTPKLTPTTPKEKAPTTTPKAPATTPPEELAPTTPEEPPTTPPEEAPT 576
Db 121 PAPTPKETAPTTPKLTPTTPKEKAPTTTPKAPATTPPEELAPTTPEEPPTTPPEEAPT 180
QY 577 TPKAAAPNTPKBPAPTTTPKBPAPTTTPKETAPTTPKETAPTTPKETAPTTLKEPAPTTPKK 636
Db 181 TPKAAAPNTPKBPAPTTTPKBPAPTTTPKETAPTTPKETAPTTPKETAPTTLKEPAPTTPKK 240
QY 637 PAPKELAPTTTKETSTSDKAPATTPKGTAPTTPKEPAPTTTPKEPAPTTTPKGTAPTTLK 696
Db 241 PAPKELAPTTTKETSTSDKAPATTPKGTAPTTPKEPAPTTTPKEPAPTTTPKGTAPTTLK 300
QY 697 EPAPTTTPKBPAPKELAPTTTKGPTSTSDKAPATTPKETAPTTPKETAPTTPKEPAPTTPKKAPTTP 756
Db 301 EPAPTTTPKBPAPKELAPTTTKGPTSTSDKAPATTPKETAPTTPKETAPTTPKEPAPTTPKKAPTTP 360
QY 757 ETDPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAETPKALENSPKPVGPTTKTPTAAT 816
Db 361 ETDPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAETPKALENSPKPVGPTTKTPTAAT 420
QY 817 KPMTTAKDKTTERDLRTPTPTTAAAPKMTKATTTTEKTTESKITATTTQVSTTTTQD 876
Db 421 KPMTTAKDKTTERDLRTPTPTTAAAPKMTKATTTTEKTTESKITATTTQVSTTTTQD 480
QY 877 TTPKITTTLKTTTLAPKVTITTKTITTTTEIMNKPEETAKPKDRATNSKATTPKQKPTKA 936
Db
```


CC indicates the presence of or predisposition to developing osteoarthritis.
CC The tribonectin and DNA encoding it are useful in the treatment of
CC osteoarthritis, where they may be used for lubricating mammalian joints,
CC such as articulating joints of humans, dogs or horses. The tribonectin,
CC when formulated as a membrane, foam, gel or fibre, is useful for
CC inhibiting adhesion between two surfaces such as the injured tissues of a
CC mammal, where the injury is caused by a surgical insertion or trauma, or
CC an artificial device e.g., an orthopaedic implant. In particular, one of
CC the surfaces is pericardial tissue. DNA encoding a tribonectin may be
CC used in gene therapy. The present sequence represents a substantial
CC portion of a human Msf-derived tribonectin
XX
XX

Query Match 51.7%; Score 2880.1; DB 4; Length 902;
Best Local Similarity 71.4%; Pred. No. 4.2e-75;
Matches 675; Conservative 27; Mismatches 74; Indels 169; Gaps 66;

QY 109 VKDNKNRRTKKKPTKPPVVDVAGSLDNGDFKVTTPDSTTQHNNKYSTSPKITTAKPIN 168
Db 1 VKDNKNRRTKKKPTKPPVVDVAGSLDNGDFKVTTPDSTTQHNNKYSTSPKITTAKPIN 60
QY 169 PRPSLPNSDTSKETSLSLVNKTETVETKTTTNNKOTSDGKEKTSKAKETOSIEKTSK 228
Db 61 PRPSLPNSDTSKETSLSLVNKTETVETKTTTNNKOTSDGKEKTSKAKETOSIEKTSK 120
QY 229 DLAPTSKVLAKPTPKAETTTKGPALTTTPKEPASTTPKEPTTTIKSAPTTPK 288
Db 121 DLAPTSKVLAKPTPKAETTTKGPALTTTPKEPASTTPKEPTTTIKSAPTTPK 172
QY 289 EPAPTTTKSAPTTKEPAPTTTKAPATTKEPAPTTTKSAPTTTKSAPTTKEPAPTT 348
Db 173 EPAPTTTKSAPTTKEPAPTTTKAPATTKEPAPTTTKSAPTTTKSAPTTKEPAPTT 228
QY 349 PKKPAETTTKEPAPTTPKAPTTPKEPAPTTKEPAPTTKEPAPTTAPKPAETTTKEPA 408
Db 229 -KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTTKEPAPTT-KEPAPTT-KEPAPTT-KEPA 281
QY 409 PTPKAPATTKEPAPTTKEPAPTTKEPAPTTKSAPTTKEPAPTTKEPAPTTKEPAPTTKE 468
Db 282 PTT-KEPAP-TPKAPATTKEPAP-TPKAPATTKEPAP-TPKAPATTKEPAP-TPKAP 335
QY 469 PAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTTKKPAETTPKAPATTKEPAPTTKEAPT 528
Db 336 PAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-TPKAPATT-KEPAPTT-KEPAPTT 388
QY 529 TPKKLPTTPTEKLAPTTPEELAPTTPEELAPTTPEELAPTTPEELAPTTPEELAPTTPE 588
Db 389 T-KEPAPTTKEP-APTKE-PAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KE 440
QY 589 PAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKE 646
Db 441 PAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTTKEPAP-T 493
QY 647 TKEPTSTSDKPAETTPKGTAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTK 706
Db 494 TKEPAPTTKE-PAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEP 546
QY 707 AP--KEPAPTTKEPAPTTSDKPAETTPKGTAPTTKEPAPTTKEPAPTTKEPAPTTKEP 764
Db 547 APTTKEPAP-TPKAPATTKEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTTKE 600
QY 765 EVSTPTTTKEPTTIHKSPESTPELSABPTPKALENSPKPEGVPPTTKTAATKPEMTT 824
Db 601 EPA--PTTKEPAPTTKEPAPTT---KEPAP---TPKAP-APTKEPAPTT----- 640
QY 825 KDKTERDLRNTTPTTAAAPKMTKETATTEKTESKITTATTQVTSSTTTQDTTPTFKITT 884
Db 641 -----TKEPAPTT-----KEPAPTTKEPAPTTKEPAP----- 667
QY 885 LKTTTLAPKVTTKKIIITTEIMNKPEETAKPKDRATNSKATTPKPKQKTPKAPKPTSK 944
Db 668 -----TPKAPATTKEP-APT-TK 684

QY 945 KKTMPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLTQTTTRPNQTPNSKLVENPKSE 1004
Db 685 EP-----APTTPTRKMTSTMPELNPTSRIAEAMLTQTTTRPNQTPNSKLVENPKSE 735
QY 1005 DAGGAGETHPHMLRPHVFPVEVTPDMDYLPVPNQGIINPMLS 1049
Db 736 DAGGAGETHPHMLRPHVFPVEVTPDMDYLPVPNQGIINPMLS 780

RESULT 14

ABUS3254
ID ABUS3254 standard; protein; 513 AA.

XX AC ABUS3254;
XX 14-APR-2003 (first entry)
XX Human testes-derived DKFZphtes3_4019 homologue #3.
XX Human; gene therapy; vaccine; disease treatment; detection.
XX Homo sapiens.
XX WO2000112659-A2.
XX 22-FEB-2001.
XX 18-AUG-2000; 2000WO-IB001496.
XX 18-AUG-1999; 99US-0149499P.
XX 28-SEP-1999; 99US-0156503P.
XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX Wiemann S;
XX WPI; 2001-327840/34.

XX Nucleic acids having the sequences of clones isolated from libraries of
XX different human tissues, useful in recombinant DNA methodologies.

XX Example III; Page 893; 1095pp; English.

XX This invention describes novel polynucleotides and polypeptides isolated
XX from human cDNA libraries which can be used for gene therapy or in
XX vaccines. The polynucleotides of the invention and antibodies encoded by
XX them may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate polypeptide expression. The products of the
XX invention may also be used to identify modulators of expression and
XX activity and to down regulate expression and activity. The antibodies of
XX the invention may also be used as diagnostic agents for detecting the
XX presence of polypeptides in samples. This sequence represents a homologue
XX of a polypeptide described in the disclosure of the invention

XX Sequence 513 AA;

Query Match 49.4%; Score 2757; DB 4; Length 513;
Best Local Similarity 100.0%; Pred. No. 7.1e-72;
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 RTKKKPTPKPPVVDVAGSLDNGDFKVTTPDSTTQHNNKYSTSPKITTAKPINRPSLPP 175
Db 1 RTKKKPTPKPPVVDVAGSLDNGDFKVTTPDSTTQHNNKYSTSPKITTAKPINRPSLPP 60
QY 176 NSDTSKETSLSLVNKTETVETKTTTNNKOTSDGKEKTSKAKETOSIEKTSKADLAPTSK 235
Db 61 NSDTSKETSLSLVNKTETVETKTTTNNKOTSDGKEKTSKAKETOSIEKTSKADLAPTSK 120
QY 236 VLAKPTPKAETTTKGPALTTTPKEPASTTPKEPTTTIKSAPTTPKEPAPTTT 295
Db 121 VLAKPTPKAETTTKGPALTTTPKEPASTTPKEPTTTIKSAPTTPKEPAPTTT 180

